

Make the most of palaeodata

Studies of past global change from sediments and ice cores have revealed a rich diversity of gradual and rapid perturbations. But more should be done to make the data usable by the research community as a whole.

One of the many depressing signals emerging from the budgetary proposals of President George W. Bush and the initial response of Congress is an absence of enthusiasm for environmental research. But there is surely no question that, as Exxon is somewhat disingenuously but rightly saying in a current advertising campaign, research into global change is essential. In the absence of reliable instrumental records spanning more than a century or so, that surely includes the study of global change over the Earth's past.

In that context, Japan, at least, can be commended for pursuing a major commitment to the next, expanded phase of the Ocean Drilling Program, starting in 2003. That programme, in conjunction with land-based drilling of ice cores and other sources of palaeo-records, has delivered a picture of the past that would have been hard to anticipate 10 years ago. Astonishingly fast excursions in global temperature and ocean circulation since the last ice age have been documented, as — further back in time — have massive emissions of methane from deep-sea clathrates, patterns of species extinction and recovery, and detailed studies of major periods of global warmth.

All of these and more have not only enlarged but also challenged our understanding of the evolution and dynamics of the Earth at its surface: the atmosphere, biosphere, oceans, ice and the mobile crust, both within themselves and also as a whole, influenced by changing fluxes in energy from the Sun caused by variations in the Earth's orbit and in solar luminosity. The orbital effects on incoming solar radiation are a vital ally for stratigraphers. For example, the 40,000-year cycle in the tilt of the Earth's axis of rotation to the plane of its orbit affects the strength of seasonal contrasts and is readily detected in sediment indicators. Understanding that and other cycles allows deep-sea sedimentologists to achieve high resolution in chronology. That helps

to disentangle influences such as changes in ocean water formation and natural greenhouse-gas abundances.

There is no alternative to committed public funding if those studying the history of the Earth's systems are to pursue these challenges and the potential they offer for the understanding of contemporary global change. These include key issues, over millennial and also much longer timescales, such as the history of sea-level changes and of the hydrological cycle, the relationships between events in the northern and southern hemispheres, and the links between the biosphere and the chemical evolution of the atmosphere.

To get the best return on such funding, there must be strong interdisciplinary collaboration between those taking data from sea sediments, ice and land, and also between those making measurements and those developing models. But above all, the community must deliver more by way of systematic deposition of its data.

In making the most of data, funding agencies, journals and researchers themselves all have a role to play. There seems to be too little awareness by researchers of what is admittedly something of a maze of publicly supported databases. And too often, as researchers will readily complain, trying to extract the numbers in a large data set from the originators of published work is like pulling teeth. That in itself is bad for science. Journals could insist on public deposition at the time of acceptance (as some do with biological data) — but only if the community had more faith in the quality of databases and their long-term support.

The latter issue points to a need for funding agencies to devote more resources to a prosaic but invaluable goal: the long-term maintenance of open archives of geoscientific data and of sophisticated tools with which to exploit them. ■

Learning to speak and write

More needs to be done to turn young scientists into comprehensible professionals and citizens.

Where does a scientist first learn the formal communication of research in talks and papers? Typically it's during doctoral training, where senior graduate students, postdocs and advisers pass on their knowledge and experiences. But this is too *ad hoc*. The remedy for the situation seems obvious: require, as part of a graduate degree, a formal course in science communication.

Even better would be to have this formal training in every undergraduate science degree, regardless of intent to pursue doctoral research. That way, scientists going into different walks of life can better communicate with non-scientists, such as work colleagues, the general public and the media. The ability to effectively communicate one's ideas and thoughts is becoming increasingly important as society and economics grow more intertwined with science and technology.

Such a proposal is hardly new, but is still all too apt. So who should be responsible for teaching at least the more formal aspects of scientific communication? In the United States, enterprising undergraduates have seized an opportunity. The National Science Foundation (NSF)

has funded a journal run by undergraduates for undergraduates that has the public support of the NSF's Rita Colwell (see page 13). Submitting to and working for this journal have been positive experiences for those who know about this resource, but it does not satisfy the wider needs of all undergraduate scientists.

The Division of Undergraduate Education at the NSF has several programmes for enhancing undergraduate science literacy and communications skills. But there is no programme that specifically aims to improve science communication for science majors. Perhaps we should look to a model developed in the physics and astronomy department at University College London, where all degree students are trained in essay writing, conducting research and giving presentations during each year of their study.

There is surely a need to spread this type of programme more widely. Anything to spare us all from more conference talks with no take-home message aided by wordy slides in tiny typefaces presented by tongue-tied, mumbling scientists. ■





Pasteur reforms
Veritable Paris
institute faces
transformation
p4



Unplugged
Researcher silenced
in digital-music
dispute
p5



Falling star
Lack of funds puts
telescope on death
row
p6



Say cheese
Three strains of
mice get the Celera
treatment
p8

Study says authors are averse to declaring conflicts of interest

Jonathan Knight, San Francisco

Authors are ignoring journal policies that require them to state any conflict of interest over their papers, a new study suggests. But editors of leading journals disagree about whether this poses a real problem and, if it does, what to do about it.

The study, from Tufts University in Massachusetts, examined 61,134 articles that appeared in 1997 in 181 leading journals that had disclosure policies. The journals examined included *Science*, *The Journal of the American Medical Association* and *Proceedings of the National Academy of Sciences*. But *Nature* and *Cell*, for example, did not have disclosure policies, and so were excluded.

The researchers counted every occurrence of a positive disclosure, including honoraria, patents pending, stock holdings or other forms of personal or financial interest. The research appears in *Science and Engineering Ethics* (7, 205–218; 2001). Only a third of the journals surveyed contained papers carrying any disclosure. Of those, fewer than 1% of papers contained a disclosure.

A possible explanation is that few researchers had anything to disclose. But the study's lead author, Sheldon Krimsky of Tufts' Department of Urban and Environmental Policy, says this is unlikely. "What is the chance that in two-thirds of the journals there was no one with a patent, equity interest or honorarium?" he asks.

But some journal editors say that such information is of little value to readers. Kevin Davies, editor-in-chief of Cell Press, says editors usually consider that good science stands on its own merits. "It's the quality of the research that counts," he says.

Science's editor-in-chief Donald Kennedy agrees. But the journal's authors must provide information on potential conflicts of interest. These disclosures are reviewed by editors, but not given to external reviewers. "The peer reviewers' job is to evaluate the science

Limited disclosure:
conflicts are rarely declared.

and not to be our ethicists," he says.

But Krimsky argues that complete



Bills threaten total US ban on human cloning

Meredith Wadman, Washington

Republicans in the US Congress have proposed two far-reaching bills that would outlaw the cloning of human cells, regardless of whether they were to be used in research or reproduction.

Senator Sam Brownback (Republican, Kansas) and Representative Dave Weldon (Republican, Florida) introduced the Human Cloning Prohibition Act of 2001 on 26 April. It bans both publicly and privately funded human cloning based on somatic-cell nuclear transfer, the technique used to produce Dolly the sheep. It also bans importation of any "product of human cloning for any purpose".

Those breaking the law would face a fine of at least \$1 million, a prison sentence of up to 10 years, or both. Another bill, introduced on the same day by Vernon Ehlers



Brownback:
sees no need for
the technique.

(Republican, Michigan), one of a handful of scientists in the Congress, also criminalizes human cloning for any purpose.

Scientists hope that cloned embryos might serve as a source of stem cells for treating a range of diseases, including diabetes and Parkinson's. But harvesting the cells means destroying the embryo, which has generated

objections from opponents of abortion. In theory, the embryos could also be used to generate a cloned human being. Already a team of infertility doctors and a cult called the Raelians have announced that they will attempt to produce cloned human beings in

the near future (see *Nature* 410, 617; 2001).

"There is no need for this technology to ever be used with humans — whether for reproductive purposes or for destructive research purposes," Brownback said in introducing his bill.

Anti-abortion groups have embraced the bill. But scientists are responding with concern. It "unnecessarily limits important medical research," says Elizabeth Marincola, the executive director of the American Society for Cell Biology. "Such a sweeping bill is not necessary" to protect against cloning a human being, she says.

The extent to which cloning should be outlawed is dividing Republicans in the Congress. Moderate Republicans are expected to back a more narrowly written bill that would preserve the cloning of embryos for research.

disclosure would help to ensure fair review. "If the article is on a safe cigarette, it would help the reviewer to know if that person had been in the public sector or employed by a cigarette company for 20 years," he says.

The *Proceedings of the National Academy of Sciences* requires contributors to disclose relevant commercial ties. But not all authors respond to the requirement without prompting, according to editor-in-chief Nicholas Cozzarelli. The journal has on occasion published corrections noting an undisclosed conflict of interest after vigilant readers complained.

Cozzarelli says the problem is one of differing interpretations of what constitutes a conflict of interest rather than any intent by authors to hide the truth.

Nature's lack of disclosure requirement is set to change, according to its editor Philip Campbell. Later this year, all *Nature* journals will ask contributors to declare any financial interest in the research they submit for publication. Contributors will have the option not to declare, but this will also be noted in the paper, Campbell says.

The shift is partly in response to evidence that financial ties do matter. For example, a study in *The Journal of the American Medical Association* (282, 1453–1457; 1999) found that research funded by pharmaceutical companies was eight times more likely than publicly funded research to reach a favourable conclusion about the usefulness of a drug.

Although the evidence is not conclusive, recent debates on genetically modified crops, for example, have highlighted public concerns about a possible undermining of scientific independence, says Campbell. "Transparency about financial interests should help to minimize such concerns," he says.

Pasteur turns to biotech firms in bid to revitalize research

Sally Goodman, Paris

Stung by criticism in a recent external review, the Pasteur Institute is implementing a range of reforms, including a biotechnology 'incubator unit', aimed at sparking a more entrepreneurial spirit among its scientists.

The incubator — Pasteur BioTop — is part of a reform programme instigated last year by director-general Philippe Kourilsky (see *Nature* 405, 990; 2000). It will eventually house 10 biotechnology start-ups, each with a two-year contract with the institute.

One of France's leading centres for biomedical research, the Pasteur Institute has an annual budget of nearly FFr1 billion (US\$135 million), operating 110 research laboratories in central Paris with 2,500 staff. Income from patents and other industrial sources cover about 40% of its costs, but it faces budgetary pressure as a number of its most important patents are due to expire.

André Choulika, head of Collectis, a genomics engineering business and one of three start-ups already housed in the incubator, says that the project has given a new lease of life to the institute.

Philip Avner, a genome researcher at the institute, agrees. "Now people who would never have had the opportunity to collaborate can have a go, and this has given a certain dynamic to the institute," he says.

But not all researchers are thrilled by the introduction of private businesses to the institute. "The Pasteur Institute risks losing its soul and originality by following fashion," said one, who declined to be identified.

Other elements of Kourilsky's reforms include changes in departmental structure and a major recruitment exercise. A recent

external evaluation of the institute, led by Harold Varmus, former director of the US National Institutes of Health, was critical of the existing structure. It said that a lack of coherence in the research areas handled by the different departments was stifling emerging talent at the institute. The report also criticized shortcomings in the Pasteur's recruitment, training and evaluation policies.

A new structure will be implemented next year, with a larger number of more tightly focused departments. Cross-disciplinary programmes are also being developed, and around 50 new research groups will be created over the next 10 years, led by young researchers recruited after an international call for proposals.

As part of the plan to attract new blood to the institute, a scholarship scheme to recruit and fund postdocs for specific research projects is also being planned.

♦ <http://www.pasteur.fr>



All change: the Pasteur Institute is attempting to capture a more entrepreneurial spirit.

INSTITUT PASTEUR

Congress hints at brighter outlook for physical science

Irwin Goodwin, Washington

Members of the Congress are backing scientists in their demands for more research funding for 2002 than has been proposed by President George W. Bush — particularly in the physical sciences.

At a meeting of the House of Representatives Science Committee on 25 April with the heads of four research agencies, chairman Sherwood Boehlert (Republican, New York) called next year's budget "particularly disappointing". He said the administration had already "signalled repeatedly that the numbers will be better next year, particularly for the National Science Foundation" (NSF), but added that

researchers should not have to wait until then for better funding.

Boehlert and other Republican committee members, in particular Connie Morella (Maryland) and Vernon Ehlers (Michigan), said they were pressing their party leaders in Congress and the White House to raise the budgets for the science agencies.

The committee asked the agency officials present, who included James Decker of the Department of Energy, the NSF's Rita Colwell and Dan Goldin of NASA, to identify programmes that would be cut or reduced to meet the budget proposal, but they declined to do so. Boehlert said he understood their reluctance, but later

described the budget proposal as "grossly inadequate".

Various committees in Congress are now considering next year's funding levels for research and other spending priorities. They will agree a final budget with Bush around the start of the 2002 fiscal year on 1 October.

In the Senate, the Budget Committee's chairman Pete Domenici (Republican, New Mexico) has complained of nearly flat funding for the energy department's science office.

An amendment to the Senate's budget resolution, calling for \$1.4 billion more science funding at the energy department, NASA and NSF, was passed last month.

Dispute over digital music muzzles academic



Corie Lok, Washington

A group of computer scientists has cancelled a conference presentation of a paper describing how they disabled anti-piracy systems for digital music, saying that they fear legal action by the recording industry.

Edward Felten, of Princeton University, and his colleagues withdrew their paper, which they had been due to present last week at the Fourth International Information Hiding Workshop in Pittsburgh, Pennsylvania, to avoid potential litigation. Felten said that he made the decision after failing to reach an agreement with representatives of the US recording industry.

"We remain committed to free speech and to the value of scientific debate to our country and the world," said Felten in a statement on 26 April, the day he was due to present the paper. "We will continue to fight for these values and for the right to publish our paper."

Later that day, the Secure Digital Music Initiative (SDMI), a consortium of recording and electronics companies that was co-founded by the Recording Industry Association of America (RIAA), released a statement saying that it did not and does not intend to bring legal action against Felten or his colleagues.

The dispute highlights the conflict between academic computer scientists, who claim that open investigation of weaknesses in security technology is necessary to improve protection in the long run, and the recording industry, which is trying to keep hackers from using this information to distribute copyrighted material.

In an attempt to detect possible shortcomings in copyright-protection measures, the SDMI challenged the public last year to crack these technologies, at least one of which was commercially available. Felten and his colleagues entered the contest, although Felten says they pulled out early and surrendered any claim on the cash prize in order to publish their findings.

Their paper describes how they successfully defeated the 'watermarks' — signals that are encoded into copyrighted digital music to prevent unauthorized copying. Called 'Reading between the lines: lessons from the SDMI challenge', it was published without permission on a libertarian website (<http://www.cryptome.org>) a few days before the conference.

Another paper detailing one of the technologies from the SDMI contest was presented as scheduled at the conference by two French researchers.

In a letter to Felten earlier this month — which also appeared on the Internet with the paper — Matthew Oppenheim, SDMI secretary and vice president of the RIAA, wrote that public disclosure of information gained from the contest could lead to unauthorized hacking of these technologies and would violate both the agreement signed by contest participants and possibly the 1998 Digital Millennium Copyright Act (DMCA).

Felten said in an e-mail, however, that he and his colleagues do not believe that their paper violates the agreement.

"The DMCA raises some serious First Amendment issues," says Lee Tien, an attorney with San Francisco-based civil liberties group the Electronic Frontier Foundation, referring to the section of the US Constitution that guarantees freedom of speech.

The DMCA outlaws public availability of

technology or devices designed to circumvent copyright-protection systems. "It is clear now that the DMCA is a serious impediment to computer-security research," says Felten, who declined to say whether he thought the release of his paper would have violated the DMCA.

Some computer scientists worry about the implications of the episode for their academic liberty. "I think to forbid the talk... is a violation of academic freedom and does not



Show-stopper: Felten withdrew his planned talk.

bring us forward to finding improvements and better algorithms," says Jana Dittmann, a watermarking researcher at the German National Research Center for Information Technology in Darmstadt.

Computer scientists' organizations have spoken out against the DMCA, but a spokesperson for the Computer Society of the Institute of Electronics and Electrical Engineers declined to comment on the Felten dispute, saying it was a political matter.

Felten did not discuss his next step in trying to publish his paper formally. However, Steven Schultz, a spokesman for Princeton University, said: "Our researchers should be able to present their research to their colleagues and the world."

Electronic ink for current issues

Declan Butler

Imagine a copy of *Nature* that looks and feels just like today's printed issue but has a cover and pages that automatically change into the latest edition. At the touch of a stylus, the magazine could be transformed into any back issue stored in an archive in the spine.

This science-fiction scenario has moved a step closer with the announcement by E Ink, a company based in Cambridge, Massachusetts, that it has developed its first portable-computer-type display — a prototype of 'electronic paper'.

E Ink's technology consists of an ink made of tiny capsules that change colour when a small electric current is passed through them (see *Nature* 394, 253–255; 1998). Coat this electronic ink onto paper, and the sheet can produce high-resolution text and images that remain visible when the electricity is switched off.

Simple electronic-ink displays have already been produced, showing only a handful of characters. To display large volumes of text and images, the technology

requires dense circuitry known as a backplane. Commercial development is limited by the lack of low-cost backplanes, says Barrett Comiskey, chief architect and co-founder of E Ink.

But the prototype display, developed by E Ink in collaboration with IBM, uses a new form of electronic ink to give the resolution of a laptop display. Because it looks like ink on paper, the screen is up to six times brighter than current laptop screens, and according to E Ink has better contrast than a newspaper, even in bright sunlight. Because it needs no backlight, the display uses one-thousandth of the power required by most portable screens. Details of the design will be presented by IBM and E Ink at the Society for Information Display Conference in San Jose, California in June.

In a step towards the ultimate goal of producing electronic paper, E Ink and Lucent Technologies have also published work (*Proc. Natl Acad. Sci. USA* 98, 4835–4840; 2001) showing the potential of using lightweight plastic backplanes.

Funding crisis could spell the end for millimetre telescope

Rex Dalton, San Diego

The National Science Foundation (NSF) has announced that it will no longer fund operation of the 12-metre telescope at Kitt Peak in Arizona — raising the likelihood that the historic instrument will finally shut down.

The funding agency has told the University of Arizona and the University of Massachusetts at Amherst that their proposal to operate the telescope for \$2.4 million for three years was being rejected because of a tightening NSF budget.

Owned by the NSF since it opened in 1967, the telescope is the only facility in the United States for millimetre-wavelength radio astronomy that is fully open to outside researchers.

The University of Arizona has been operating the telescope since last July, when the NSF said that it would no longer bear the operation costs on a permanent basis. The Tucson-based Research Corporation stepped in with philanthropic funds for interim operations while a new NSF grant was sought.

Reviewers of the proposal to the NSF gave the project high marks, Arizona officials say, but NSF administrators said money was not available given the Bush administration's budget proposal.

"We are simply dumbfounded," says Peter Strittmatter, director of Arizona's Steward Observatory and leader of the grant proposal to the NSF. "The importance of the telescope to the US astronomy community was clearly recognized by all the reviewers."

The University of Arizona will now begin a search for alternative funding to try to keep the telescope open for the 100-plus astronomers that use it. "We will keep our shoestring operation going" while trying to find groups to guarantee purchase of time on the telescope, says Strittmatter. ■



On the brink: proposals to keep Kitt Peak's 12-metre telescope open have been rejected.

Professors facing power cuts in German university reforms

Marco Jäger and Quirin Schiermeier, Munich

German universities are set for a radical reform that will curb the pervasive powers of senior professors and provide more freedom and financial incentive for junior ones.

Under legislation presented last week by science minister Edelgard Bulmahn, professors will be paid on the basis of performance rather than seniority. But the rules, which should be in force by next year, will only apply to new professors and to existing ones who choose to take them up.

The reforms are opposed by academic organizations in Germany, despite the fact that many younger academics support the change.

Under the proposals, the archaic qualification required for professorship in Germany — known as *Habilitation* — would be phased out by 2010. Junior professorships, similar to US assistant professorships, would gradually replace *Habilitation*.

In addition, professors' basic monthly salaries, which currently range from DM7,000 (US\$3,300) to DM18,000 depending on age and position, would be reduced by a third, and bonuses would be paid to those who meet or exceed performance targets. The current DM18,000 salary limit would be lifted — a move that is considered necessary to attract top international researchers to German universities.

Bulmahn last week sent the reform proposals to the 16 *Länder* (state) governments, which are responsible for education. They are expected to enact the changes as they were involved in the commission that drew up the reforms.

But Germany's academic establishment has already criticized the changes, which have been under discussion for two years. Last March, for example, almost 4,000 faculty members, including emeritus professors, signed a letter of protest published as an advertisement in the newspaper *Frankfurter Allgemeine Zeitung*.

The advertisement was paid for by the Deutsche Hochschulverband, the largest association of German university professors. Hartmut Schiedermaier, president of the association, strongly opposes Bulmahn's plans.

Some younger scientists disagree with the association's position. Hartmut Engelmann, for example, a 42-year-old immunologist at the University of Munich, who underwent *Habilitation* two years ago, says he will be glad to see the back of the process.

"There is no transparency at all, only personal dependence," he says. "Everything is fine if your institute head is well-disposed



In with the new: science minister Edelgard Bulmahn aims to encourage younger academics.

towards you. If not, it can eat you up."

Engelmann also welcomes the idea that professors should be paid on the basis of the quality of their teaching and research. But he admits that it will be difficult to carry out fair evaluations.

Bulmahn has not yet specified details of how these evaluations will be done. According to a spokeswoman for the education department, options include assessing the number of scientific publications and using students' evaluations of lectures. "Only practical experience will teach us where to position the lever," she says. ■



Standing firm: Hartmut Schiedermaier is leading professorial opposition to the reforms.

Researchers strike back in animal-rights row

David Adam, London

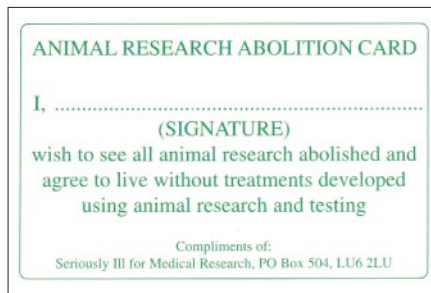
A group of charities is to move its bank account following its bank's refusal to guarantee that it would stand up to public protests against animal research.

The Association of Medical Research Charities (AMRC) said on 23 April that it would close its account with the HSBC bank after the HSBC refused to confirm that it would stand by customers if they were targeted by animal-rights activists.

The HSBC was criticized recently for bowing to animal-rights campaigners when it joined other banks and stockbrokers in refusing to handle shares in the drug-testing group Huntingdon Life Sciences (HLS) following threats from activists.

The AMRC's move is seen as part of a trend towards more aggressive tactics by British medical researchers in their public-relations war against animal-rights activists.

For example, the BioIndustry Association, which represents the UK biotechnology industry, wrote to venture capitalists in March urging them not to cave in to antivivisection protesters. And in a joint letter to *The Times* newspaper last week, the heads of a dozen academic and commercial research organizations pointed to the "huge contribution that laboratory animal research has



Making a point: animal-rights activists might try carrying this card, their opponents suggest.

made to medical science". The patients' group Seriously Ill for Medical Research has sent spoof donor cards to animal-rights protesters asking them to refuse medical treatment.

An AMRC statement said: "Following recent targeting by animal rights activists of banks and investment funds, AMRC had sought assurances from HSBC that the bank would not give in to a similar campaign... AMRC executive council considered HSBC's responses unsatisfactory and has agreed to move its account."

The AMRC has a few hundred thousand pounds in its account, but its 112 members, which include the Wellcome Trust and the Cancer Research Campaign, have combined assets of almost £16 billion (US\$23 billion).

To honour my belief that animal research should be abolished, I hereby pledge that:

- in the event of accident or emergency, I will refuse all medical treatments developed or tested on animals, including, but not limited to: blood transfusions, anaesthetics, anticoagulants, antibiotics, sutures, open heart and other types of surgery.
- if my child suffers from a genetic illness or other serious condition, I will not allow them to have life-saving treatment developed through animal research.
- none of my pets shall receive any veterinary vaccine or medicine that has been developed or tested on animals.

Diana Garnham, the AMRC's chief executive, says she expects many of its charities to seek similar assurances from their banks. The Wellcome Trust says it is "keeping a careful eye on how events unfold".

A spokesman for HSBC says the bank was "disappointed" with the AMRC's decision. The association put the bank in an impossible position, he says, by demanding a "100% legal guarantee" to stand by its members, whatever the circumstances.

The campaign against HLS included break-ins at its suppliers' factories and threats to its investors. In February, the company's managing director was attacked by masked assailants wielding baseball bats (see *Nature* **410**, 8; 2001).

Fungus fingered as dogwood disappears from forests

Mark Schroppe

Flowering dogwoods, *Cornus florida*, whose white blossoms were once a welcome and common sign of spring's arrival in the United States east of the Mississippi, are rapidly disappearing.

The loss is mainly the result of an exotic fungus called anthracnose (*Discula destructiva*), which causes lesions on leaves and trunks, and eventually kills the plant. The fungus was probably introduced into the United States decades ago on nursery trees from Asia, and was first spotted on the east coast in around 1976 before spreading rapidly throughout eastern states.

In a recent study in the Great Smoky Mountains National Park in Tennessee and North Carolina, for example, National Park Service forest ecologist Michael Jenkins and colleagues have found evidence of severe declines in the flowering dogwood population since the 1970s in a variety of forest habitats. Similarly dismal results have also been found in other states.

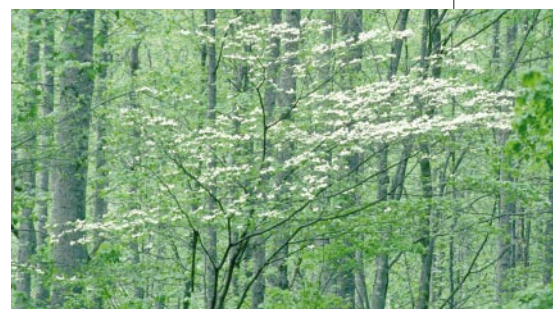
Peter White, a plant ecologist at the University of North Carolina in Chapel Hill and an author of the Great Smoky Mountains study, says that, although the

fungus seems to be the cause of the population decline, other factors may have contributed to its spread. For instance, the end of older Native American practices such as intentional burning means that fire is now far less prevalent. This results in thicker forest canopy and increased moisture, which helps the fungus to thrive.

The dogwood population has also declined in Indiana's Ross Biological Reserve, but not as severely as in the Smoky Mountains. Kerry Rabenold, a bird ecologist at Purdue University in West Lafayette, Indiana, says the reserve has also experienced changes that would help the fungus to spread, such as less fire. But anthracnose has not been found there.

Changes in the make-up of the forest appear to be affecting the reserve's dogwood population. Less fire helps maple trees to thrive, forming a denser canopy that stops dogwoods getting the amount of sunlight they need. Rabenold says that other understory trees not susceptible to the fungus have also been dying off.

Dogwoods are an important part of the forest ecosystem. They are a source of calcium for animals and plants, for example,



Blooming lovely: but the flowering dogwood is in decline on the Great Smoky Mountains.

drawing it from deep soil and concentrating it in their leaves.

The trees also produce berries, which Rabenold says are "probably the best food in North America for songbirds". He believes that declines in dogwood have adversely affected bird populations.

White says effects similar to those cited in the Indiana work are probably at play in other areas with dwindling dogwood populations, but that, in his study and many others, anthracnose is the clear culprit. "You might compare it to being run over by truck versus a chronic disease," he says.

CORBIS

Celera completes genome sequences of three mouse strains

Washington Celera Genomics has announced that it has sequenced the genomes of three strains of mice. The company says that the assembled sequences each contain about 2.6 billion base pairs, compared with 2.9 billion base pairs in the human genome.

Celera, based in Rockville, Maryland, this year published a draft sequence of the human genome (see *Science* **291**, 1304–1351; 2001). It aims to capitalize on its latest investment by charging scientists to use the mouse genome and the tools needed to compare it with the human genome and those of other organisms.



Shotgun sequence: Celera will ask researchers to pay for access to its three mouse genomes.

The mouse genome is of particular interest because mice are used as models for many human diseases. A publicly funded group currently sequencing the genome of another mouse strain hopes to have a draft by 2003.

Oxford and Princeton forge stronger links

London The universities of Oxford and Princeton have announced plans to improve ties between the two institutions. The scheme will create new research partnerships, increase faculty and student exchanges, and give researchers at each institution access to facilities at the other.

The programme will initially cover 12 collaborative projects, encompassing research in nanotechnology, astrophysics, materials science, biotechnology and genomics. The universities are “remarkably similar in goals and strengths”, says Oxford vice-chancellor Colin Lucas. “This agreement will help make both institutions even stronger.”

Foot-and-mouth spurs calls for safe soil

London Fears that carcasses of animals slaughtered during Britain’s foot-and-mouth epidemic could be spreading pathogens to

farmland and drinking water have led scientists to call for stronger protection of the soil.

“We need laws to ensure that disposal is as rapid as possible and always in the right location,” says Ken Killham, president of the British Society for Soil Science. “Things must happen now, we mustn’t let the continued unsustainable practices in planning and the disposal of carcasses ever happen again.”

The society says that recent government proposals for a national soil strategy are “disappointing” because they do not include legislative protection for soil. It wants the soil to be given similar protection to that currently granted to air and water.

♦ <http://www.bsos.bangor.ac.uk>

Vandals cut power to Arizona observatory

San Diego The operators of Mount Graham International Observatory in Arizona have blamed extreme environmentalists for causing damage worth more than \$100,000 to the construction of a new power line to the observatory.

Some environmentalists object to the expansion of the telescope site, of which the new 37-kilometre underground power line is part. They claim that nearby populations of

red squirrels and spruce trees would be harmed if scientists were allowed to go ahead with proposed plans for new telescopes at the observatory. Work on the \$10.5-million project was suspended for a day after vandals damaged vehicles and equipment.

Environmental groups had earlier failed in attempts to persuade the courts to block the observatory's expansion.

MMR vaccine cleared of link with autism

Washington There is no evidence for a link between the measles–mumps–rubella (MMR) vaccine and autism, a study from the US Institute of Medicine reports.

A paper published in *The Lancet* in 1998 had described how 12 children had developed autistic spectrum disorders shortly after being given the MMR jab. That research has since been heavily criticized, and the new report is the latest in a series of epidemiological studies that have failed to confirm the link. But its authors point out that almost all US children receive the MMR jab, making it difficult to compare vaccinated and non-vaccinated populations.

Public-health officials have expressed concern that publicity surrounding the *Lancet* paper would dissuade parents from

having their children vaccinated. The MMR vaccine has helped to virtually eliminate measles, mumps and rubella in the United States.

► <http://www.nap.edu/books/0309074479/html>

Three scientists appointed to House of Lords

London Three scientists have been chosen to join the British House of Lords under a new scheme to appoint independent life peers.

Robert May, president of the Royal Society, Susan Greenfield, director of the Royal Institution, a London-based research and education organization, and Ilora Finlay, vice-dean of the University of Wales College of Medicine, will take their seats after this summer's general election. The election committee said the new peers were chosen for their "knowledge, authority and independence".

Twelve other appointments were made under the so-called people's peers scheme,



Life peer: Susan Greenfield will take her seat in the House of Lords after the general election.

which has been criticized for not selecting ordinary members of the public.

Appointment of the people's peers is the latest stage in the reform of the House of Lords following the expulsion of 600 hereditary peers last year. The government is yet to decide how the remaining positions will be filled.

New site licence

The publishers of *Nature* are pleased to announce the introduction of new terms of institutional site licences for *Nature* online, to apply from 1 May 2001. From that date, existing and new holders of site licences will have access to all of *Nature*'s content at the date of publication, under terms described at the *Nature* website.

► <http://www.nature.com/help/sitelicences>

On the trail of the neutrino

Huge arrays of detectors now have these ghostly particles in their sights — but will what they see lead physicists to rethink the standard model? Dan Falk investigates.

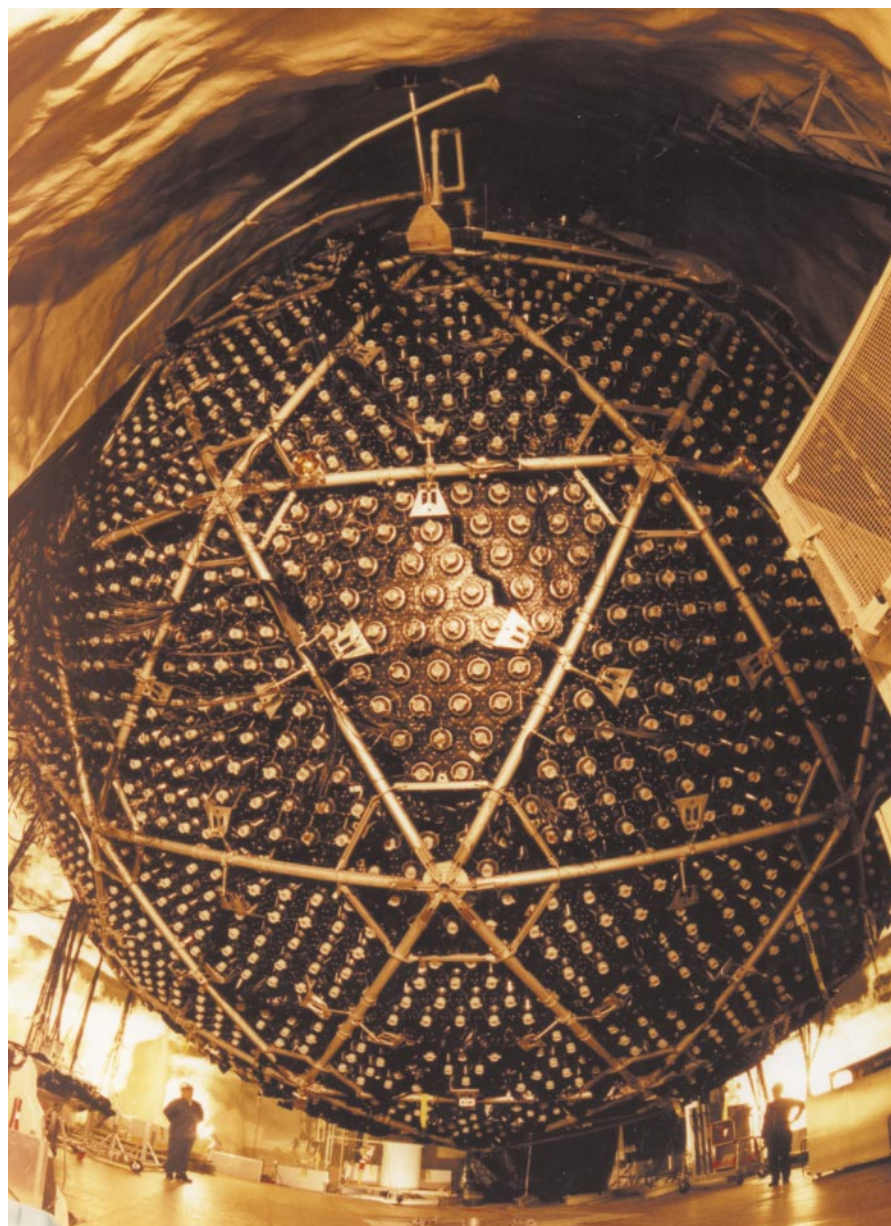
In the subatomic zoo, few particles are as elusive as the neutrino. The Universe is awash with them, but they slide through most forms of matter with ease. Billions have passed through your body since you started reading this article. By one estimate, the average neutrino could travel for 1,000 light-years through solid matter before being stopped.

This reluctance to interact with matter makes neutrinos difficult to detect. But, almost half a century after they were first spotted, neutrinos have become the focus of intense study. Exciting results from existing detectors, together with plans for new detector projects, promise to make this a busy decade for neutrino hunters.

The Sudbury Neutrino Observatory (SNO) in Ontario is poised to illuminate the physics of neutrinos produced by the Sun, and a new detector in Antarctica is making progress towards studying neutrinos from deep space — a first step towards building a larger astrophysical neutrino observatory. Other detectors in North America, Europe and Japan are catching neutrinos produced when cosmic rays hit the atmosphere or by particle accelerators here on Earth. “I think it’s fair to say that this is the era of the neutrino,” says John Bahcall of the Institute for Advanced Study in Princeton, New Jersey, a pioneer of neutrino physics.

Three of a kind

According to the standard model, the successful but incomplete theory that describes fundamental particles and their interactions, neutrinos come in three types or ‘flavours’: electron, muon and tau neutrinos. The names derive from links within the model between the neutrinos and other particles — the familiar electron and its rarer, heavier cousins the muon and tau



Full house: the detector at Sudbury Neutrino Observatory can detect all three flavours of neutrino.

particles. The last of the three flavours to be detected was the tau neutrino, finally snared last year in an experiment at the Tevatron accelerator in Fermilab, near Chicago¹.

The problem faced by neutrino hunters is that all three flavours of neutrinos are extremely unwilling to interact with other subatomic particles. Of the four types of fundamental interaction between particles, neutrinos interact almost exclusively through the weak nuclear force, the puny force responsible for some radioactive decays. As they do not carry charge, neutrinos avoid electromagnetic interactions. Nor do they interact through the strong nuclear force that binds nuclei together. The jury is still out on whether neutrinos feel gravity, but if they do, the forces involved will be tiny.

So the only practical way to snag a neutrino hinges on the weak nuclear force and the interactions between neutrinos and protons,

neutrons or electrons it governs. Physicists maximize their chances of observing these rare interactions by monitoring huge numbers of potential neutrino targets. Large tanks of water are a common choice — water is cheap, plentiful and, being transparent, it allows physicists to spot any interactions that occur deep inside the tank. The tanks are held deep below ground, so that other particles are screened out by the overlying rock.

There are several types of interaction to look for. The simplest is ‘elastic scattering’ — high-speed collisions between electrons in water molecules and incoming neutrinos. The collision leaves the electrons travelling faster than the speed of light through water. As a result, the electrons emit dim, blue flashes of light known as Cerenkov radiation — the optical equivalent of the ‘sonic boom’ created by supersonic aircraft. Neutrinos can also collide with neutrons or protons in

atomic nuclei, and some of the particles produced in these reactions also travel fast enough to produce Cerenkov radiation. In each case, the flashes of light can be captured by photomultiplier tubes (PMTs) on the sides of the detector tanks.

Detectors such as these could help to settle the most pressing issue in neutrino physics — that of the neutrino mass. The standard model says that neutrinos should have zero mass, but evidence from solar physics and some proposed Grand Unified Theories — efforts to unify electromagnetism and the strong and weak nuclear forces as different aspects of a single interaction — points instead to the particles having a small mass. For physicists, the tell-tale sign that neutrinos have mass would be ‘flavour-switching’, or oscillation. Neutrinos oscillate if they switch from one flavour to another as they travel. The way in which such switching would occur is not fully understood, but physicists know that to oscillate neutrinos need to have mass².

In 1998, a team of US and Japanese researchers using the Super-Kamiokande, or Super-K, detector in Japan found evidence that appeared to support the idea of neutrinos with mass³. The story landed on the front page of *The New York Times* under a headline declaring that the “Universe May Never Be the Same”.

Something in the air

Buried deep under Mount Ikeno near the town of Kamioka, 200 kilometres northwest of Tokyo, the Super-K detector uses 13,000 PMTs to monitor a 50-million-litre tank of water. It searches for signatures of ‘atmospheric neutrinos’ — electron or muon neutrinos produced when high-speed atomic nuclei and electrons from deep space collide with oxygen or nitrogen atoms in the Earth’s atmosphere. These neutrinos can be singled out because they come from a different direction and have a different energy compared with other neutrinos.

The present theory of how atmospheric neutrinos are generated suggests that Super-K should detect twice as many muon neutrinos as electron neutrinos. Instead, it is recording roughly equal numbers of each³. This, argues the Super-K team, shows that some of the muon neutrinos are turning into something else — tau neutrinos — as they travel through the atmosphere and the Earth’s crust. Super-K cannot detect tau neutrinos because the particles produced when they collide with atomic nuclei are usually too short-lived to generate Cerenkov radiation.

Although the Super-K results cannot give a specific value for the mass of any neutrino flavour, they can be used to calculate a mass difference — in this case, a fraction of a millionth of the mass of an electron — between the muon and tau types. In other words, either the muon or the tau neutrino must have a mass at least this large.

Firm data on neutrino mass could point the way to a more elegant theory.

But the experiment is not foolproof. The calculation depends on the rate at which atmospheric neutrinos are produced, and this is not known precisely. Those involved with the project are confident that Super-K has indeed seen neutrino flavour-switching. “The oscillation theory quite nicely explains our data,” says Yoji Totsuka, the observatory’s director and head of the Institute for Cosmic Ray Research at the University of Tokyo.

Solar flaws

Conclusive evidence for neutrino oscillation would also solve one of the longest-standing questions in solar physics. From what we understand about how the Sun produces neutrinos, 60 billion solar neutrinos should pass through every square centimetre of the Earth’s surface every second. But measurements from neutrino detectors suggest that only between a half and a third of this figure are reaching the Earth^{2,4}.

Either the models that describe the Sun’s interior are flawed — and most scientists doubt this to be the case — or something is happening to the neutrinos between their emission from the Sun’s core and their arrival on Earth. In the main, the Sun generates electron neutrinos, but most physicists think that many of these switch flavours en route, turning into muon neutrinos or tau neutrinos. Neutrino detectors, including

Super-K, have studied the problem, but because they are not equipped to distinguish between the various solar-generated neutrinos, the results have been inconclusive.

The SNO, a joint project between Canada, the United States and Britain, has perhaps the greatest chance of resolving the issue. Located at the bottom of one of North America’s deepest mines, 300 kilometres north of Toronto, the SNO’s detector is a spherical vat 12 metres across, filled with 1,000 tonnes of heavy water and lined with 10,000 PMTs. Heavy water contains deuterium — an isotope of hydrogen that has a neutron as well as a proton in its nucleus.

Most Cerenkov radiation observed at other detectors comes from interactions between electron neutrinos and either neutrons or electrons. But the SNO’s deuterium atoms allow it to detect another type of collision which is more sensitive to the other neutrino flavours. Incoming neutrinos — electron, muon or tau — can, very occasionally, split the proton and neutron in deuterium apart. Gamma rays are generated if the neutron released by this fission is subsequently captured by another deuterium nucleus, and the SNO’s PMTs can detect this radiation.

This ability to catch all three types of neutrinos is the SNO’s big advantage. By comparing the numbers of each type of solar neutrino arriving at Earth with the number expected from theories of solar physics, the SNO should be able to settle the neutrino flavour-switching issue. The SNO team should have a major announcement within months, says Art McDonald, the observatory’s director.

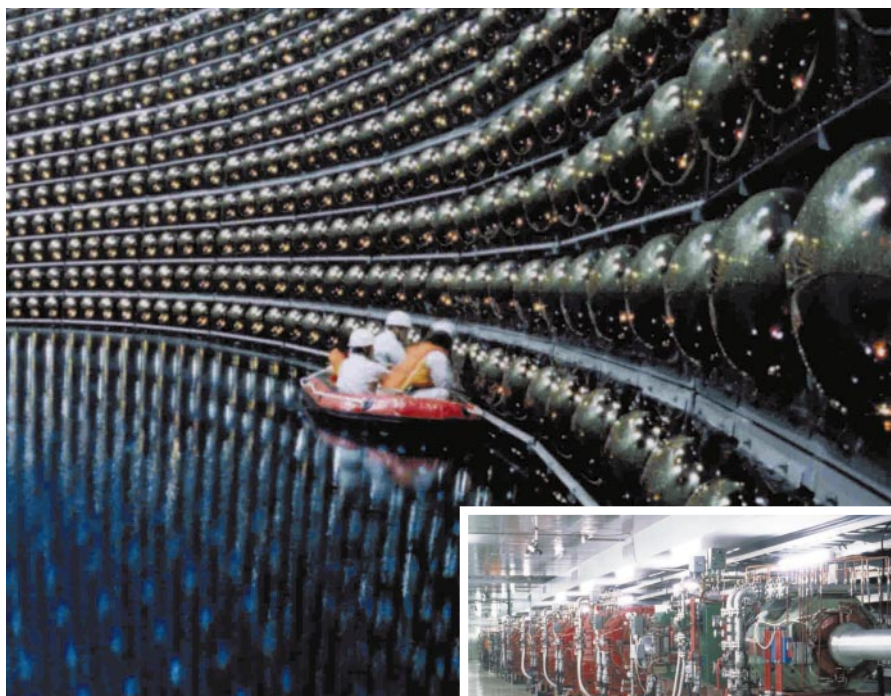
Other physicists are studying neutrinos produced by artificial sources such as nuclear reactors and particle accelerators. Researchers using these sources have more



Eyes down: One of AMANDA’s photomultipliers is lowered into the Antarctic ice. AMANDA is a precursor to a full-blown neutrino telescope.



ROBERT MORSE



Long shot: a beam of neutrinos from KEK (inset) will be fired through 250 km of the Earth's crust to Super-K (main picture).



information about how the neutrinos started out — where they came from, their energy and their total numbers. This allows them to look for oscillations under more controlled conditions.

In Japan, physicists at the KEK facility — the High Energy Accelerator Research Organization — near Tokyo are using a particle accelerator to produce a beam of muon neutrinos. This is aimed through the Earth's crust towards Super-K, 250 km to the west. Physicists can estimate how many muon neutrinos Super-K should detect over long periods of time. After two years, says Totsuka, it has seen 28 neutrinos⁵. The theory says the detector should have seen 38, suggesting that some of the muon neutrinos are turning into undetectable tau neutrinos on the way. The Super-K team says that the rate of switching inferred from the beam experiments is consistent with atmospheric neutrino studies. By July, says Totsuka, the team should have doubled the amount of data collected, and will publish more definitive results within a few years.

Two other similar experiments are in the planning stage. In Europe, physicists are going to fire a beam of neutrinos from CERN, the European Laboratory for Particle Physics near Geneva, through the Alps to Italy's Gran Sasso National Laboratory near Rome, 730 km away. The project should start collecting data in 2005. Rival physicists in the United States hope that the MINOS (Main Injector Neutrino Oscillation Search) project — a 725-km neutrino beam from Fermilab to the Soudan mine in Minnesota — will start producing results a year earlier.

Another 'controlled conditions' ap-

proach will take advantage of the neutrinos produced in nuclear power plants. Later this year, a project called KamLAND (Kamioka Liquid Scintillator Antineutrino Detector) will start trying to detect antineutrinos, the antimatter equivalent to neutrinos, emitted by Japan's 51 nuclear power reactors, as well as solar neutrinos. This US-Japanese collaboration is housed in the same mine as Super-K, but uses a detector based on a liquid scintillator, a chemical that luminesces when low-energy neutrinos and antineutrinos pass through it.

The hole picture

Astronomers are also eager to get in on the neutrino detection act. They want to exploit neutrinos as a tool for studying some of the most energetic phenomena in the cosmos. As they travel through space, light and radio waves can be scattered or absorbed by gas and dust, but neutrinos carry information from their source to the Earth almost undisturbed, so astronomers can use them to study the farthest reaches of the Universe.

The dream of an operational neutrino telescope came a step closer in March, when the Antarctic Muon and Neutrino Detector Array (AMANDA) in Antarctica reported spotting its first atmospheric neutrinos⁶. Buried in the ice some two kilometres below the South Pole, AMANDA consists of nearly 700 PMTs arranged in a huge cylinder, half a kilometre high and 200 metres in diameter. AMANDA is actually 'looking' downwards, searching for neutrinos that have travelled through the Earth's core. Like the layers of rock above the subterranean detectors, this

screens out other subatomic particles. AMANDA detects fast-moving muons produced by collisions between highly energetic neutrinos and protons or neutrons. Other detectors miss the Cerenkov radiation produced by these muons because they move so quickly that they leave the detector before emitting useful amounts of light, but AMANDA's huge size allows it to spot them.

Innovative as AMANDA is, it is really only a stepping-stone towards a much larger cosmic neutrino detector, says Francis Halzen, part of the University of Wisconsin-Madison team that is leading the collaboration of European and US institutions behind the project. The partners are already drawing up plans for a much larger detector, dubbed IceCube. This should be able to detect neutrinos from cosmic sources such as black holes, quasars, supernovas and gamma-ray bursts. Bahcall likens AMANDA to a Model T Ford, and says: "IceCube will be the real Cadillac."

Physicists plan to use IceCube to study what happens to material approaching the event horizon of a black hole — the point of no return for the matter and light that the black hole is consuming. Clouds of hot swirling gas and dust around the black hole block the gaze of conventional telescopes, but neutrinos stream straight through, giving physicists an insight into this previously hidden region.

Addressing such mysteries should keep astrophysicists busy, but the most profound repercussions of the new era of neutrino studies could be for the standard model. If neutrinos do have mass, it will be the theory's first serious empirical challenge. "The finite neutrino mass is the first observational result beyond the standard model," says Totsuka.

If Super-K's initial results are backed up by the SNO and other detectors, the standard model will have to be corrected. And few physicists will be sorry. "The standard model is a paradox, because it's enormously successful in terms of describing a huge amount of data, but it has some features which are ugly — like having a whole bunch of arbitrary parameters," says Bahcall. Firm data on neutrino mass could help eliminate the ugliness and point the way to a more elegant theory. ■

Dan Falk is a freelance writer based in Toronto.

1. Kodama, K. *et al. Phys. Lett. B* (in the press).
2. Kirsten, T. A. *Rev. Mod. Phys.* **71**, 1213–1232 (1999).
3. Fukuda, Y. *Phys. Rev. Lett.* **81**, 1562–1567 (1998).
4. Bahcall, J. N. *Phys. Rep.* **333**, 47–62 (2000).
5. Nakamura, K. *Nucl. Phys. B — Proc. Suppl.* **91**, 203–209 (2001).
6. Andrés, E. *et al. Nature* **410**, 441–443 (2001).

Web links

Sudbury Neutrino Observatory

♦ <http://www.sno.phy.queensu.ca>

Super-Kamiokande

♦ <http://www-sk.icrr.u-tokyo.ac.jp/doc/sk>

AMANDA

♦ <http://amanda.berkeley.edu/amanda/amanda.html>

IceCube

♦ <http://pheno.physics.wisc.edu/icecube>



The youth team

Undergraduates are not only writing scientific papers — they're reviewing them, editing them and posting them online. Josette Chen examines a scientific publishing phenomenon.

Trying to get your first research paper published can be a daunting experience. After writing and revising *ad infinitum*, you finally submit to a leading journal — only to have your self-esteem crushed by a scathing referee's report. Wouldn't it have been better to hone your paper-writing skills while you were still an undergraduate, learning the tricks of the trade with help from your peers?

That is the thinking behind a crop of student journals that has emerged in the past few years. Set up to publish the results of undergraduate research projects, some of these operate solely online, avoiding the high production costs of print publications. Several are run entirely by undergraduates, with their editorial staff recruited from student ranks. And, like journals such as *Nature*, some even include news, comment and feature articles, in addition to publishing original research. Together, they are providing valuable experience in all aspects of scientific publishing for the researchers, journal editors and science writers of tomorrow.

Leading the pack is the international web-based *Journal of Young Investigators (JYI)*,

which has published three issues in the past couple of years, and is now gearing up to a regular output of three issues per year. It has an editorial staff of around 30 part-time undergraduate volunteers. Submitted manuscripts are reviewed by associate editors in conjunction with their faculty advisers. Recommendations are passed to section editors, who deal with different disciplines, and ultimately to the editor-in-chief — currently Courtney Peterson, who studies physics and biology at Georgetown University in Washington. "We try to get each manuscript through the whole process in five weeks," she says.

The *JYI* began in 1997 as a twinkle in the eyes of five undergraduates from Duke Uni-

We cooped ourselves up, eating cold Chinese food, to come up with our mission statement.

versity in Durham, North Carolina, Swarthmore College in Pennsylvania and Brown University in Providence, Rhode Island. Neal Freedman, George Lui, Andrew Medina-Marino, Tim Sibley and Brian Su felt that research by undergraduates was not getting due recognition and decided to remedy the situation. "We gathered in a Baltimore hotel and cooped ourselves up for two and a half days, eating cold Chinese food, to come up with our mission statement and budget," recalls Su, now a medical student at Columbia University in New York, but then studying biomedical and electrical engineering at Duke.

The five students reckoned it would take at least \$20,000 to launch an online journal. "We just started laughing," says Su. "This is a huge sum for undergraduates." But the idea was warmly received, and after a year, Su and his colleagues had raised around \$200,000 from the National Science Foundation (NSF), the Burroughs Wellcome Fund, Glaxo Wellcome, Duke and Swarthmore. Through mass e-mailings and campus visits, the team recruited editors and solicited submissions — eventually going live with the journal's first issue in December 1998.

Boundless enthusiasm

There have been growing pains, mostly associated with the logistics of dealing with a dispersed staff working on a volunteer basis, with a high turnover rate. And one mathematics paper, outlining a proof for the irrationality of Euler's constant, was withdrawn after a former section editor spotted errors in the published version. But with plenty of e-mail and phone contact, yearly editorial conferences, and boundless enthusiasm, many of the wrinkles have been ironed out. The journal's founders, and many former staff, are continuing to share their experiences through the *JYI*'s Board of Advisors. Peterson believes the quality of submissions is improving. And with her current staff, she is now finalizing the *JYI*'s second NSF grant application, for around \$100,000.

The journal is also expanding internationally. Andreas Bender, who is studying chemistry at the Technical University of Berlin, joined the *JYI* early last year with a mandate to add European papers and staff. He has recruited James O'Dwyer from Durham University in Britain as a section editor for physical sciences and mathematics.

Budding scientists who have published in the *JYI* are overwhelmingly positive about the experience. Elise Donnelly, who studied biology at the College of William and Mary in Williamsburg, Virginia, published a paper in the *JYI*'s first issue on nest defence in blue-headed vireos (*Vireo solitarius*), birds in which both sexes play equal roles in parenting. Now finishing a masters programme at Wake Forest University in Winston-Salem, North Carolina, and planning for a PhD at the University of California, Santa Cruz,

Fellow undergraduates are more understanding than major figures in the field.

► Donnelly says that writing her *JYI* paper helped build her skills and confidence. She is now preparing two papers for front-line journals. "People who've read the drafts say my writing is good," she says.

Brian Skotko of Duke University worked on a Howard Hughes Medical Institute undergraduate research fellowship in neuroscience, studying the effect of the nucleoside adenosine in a region of the rat hippocampus involved with processing spatial memory. After completing his final report, he submitted his work to the *JYI*. "I really got to learn science methodology, such as making hypotheses, solving problems and answering questions," says Skotko, who now plans to enter medical school.

Even though the *JYI*'s review process is tough — only around 10% of the 200-plus submissions received so far have been accepted for publication — its editors believe it provides a gentle introduction to getting a paper published. "You are being reviewed by fellow undergraduates, who are more understanding than major figures in the field," says Mary Patyten, a *JYI* associate editor.

Gains all round

The *JYI* has also provided valuable experience for its staff. Adam Friedman, now in his final year of a biology degree at Princeton University in New Jersey, says that editing other people's work has made him more critical of his own writing and how he presents his science. And Patyten, who helped develop *JYI*'s features content while studying Earth systems science and policy at California State University, Monterey Bay, is now a science writer with the California Department of Fish and Game in Marina. The journal's staff can stay on for a year after they graduate, and Patyten is using that time to run a mentoring programme for aspiring science journalists who want to contribute to the *JYI*.

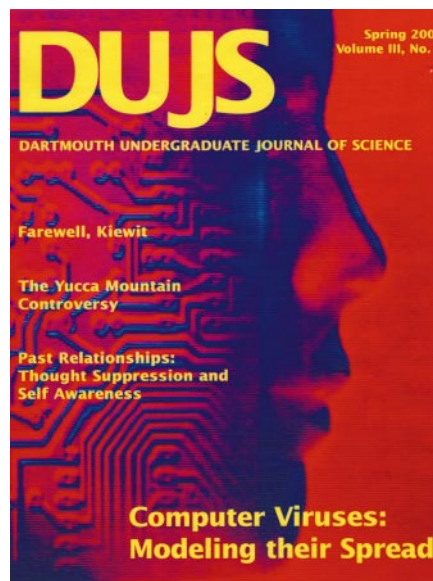
The *JYI* is just the most prominent of a string of student journals to have emerged in the past few years. Many are associated with particular universities, and publish the work of their host institutions' students. They follow in a long tradition: the *University of Toronto Medical Journal*, which is student-run, was first published in 1923. But the potential for affordable scientific publication offered by the Internet has encouraged the launch of many others.



Looking ahead: *Journal of Young Investigators* editor-in-chief Courtney Peterson believes the quality of submissions is improving.

Typical of the new breed is the *Dartmouth Undergraduate Journal of Science (DUJS)* formed in 1998 at Dartmouth College in Hanover, New Hampshire, by four industrious undergrads. In addition to the online journal, a hard copy of the *DUJS*, which features eye-catching cover designs (see below), is produced for campus-wide distribution. The review of manuscripts is conducted completely by undergraduates.

Again, students who have published in the *DUJS* say they have benefited from the experience. Anura Abeyesinghe, a physics undergraduate who published on quantum entanglement, says the paper proved valuable when he applied for graduate school. "The journal is online and that made it easy for me to simply mention the website in my



Widely read: the *DUJS* is on paper and online.



application so that the grad school people could read my article and get an idea of the kind of work I have done." He intends to study quantum information theory at either the Massachusetts Institute of Technology or the California Institute of Technology.

Other projects are subject-specific but not limited to research by students at a particular institution. The National Undergraduate Research Clearinghouse, for instance, provides an online repository for psychology papers. It was established three years ago by Brian Cronk, a professor at Missouri Western State College in St Joseph. Rather than using conventional peer review, it exercises quality control by getting faculty members to sponsor the submissions.

Encouraged by the success of current journals, others are poised to join the fray. Sinah Shah and Alex Kastor, two graduate students at the University of Pennsylvania in Philadelphia, have just launched a multi-disciplinary web-based journal, the *Journal of Undergraduate Study & Independent Research*. Faculty from across the United States have volunteered to serve as reviewers.

Sceptics might argue that an undergraduate journal is no substitute for a 'real' scientific publication. But influential figures believe that such journals are performing a useful function: the *JYI*'s last editorial conference, held in February at the annual meeting of the American Association for the Advancement of Science in San Francisco, was addressed by Rita Colwell, director of the NSF. The enthusiasm of young scientists who have published in the journals also suggests that there is a real demand among undergraduates for more experience in the art of getting your science published.

Josette Chen is an intern at Nature.

Journal of Young Investigators

► <http://www.jyi.org>

University of Toronto Medical Journal

► <http://www.utmj.org>

Dartmouth Undergraduate Journal of Science

► <http://www.dartmouth.edu/~dujs>

National Undergraduate Research Clearinghouse

► <http://clearinghouse.mwsc.edu>

Journal of Undergraduate Study & Independent Research

► <http://www.jusir.org>

Links to other undergraduate journals

► <http://eden.mercy.edu/upd>

► <http://www.jyi.org/resources/ugradPubs.html>

Who will fill the gap by making nucleic synthesizers now?

Sir — We would like to draw attention to a looming problem for all those of us doing research involving the chemical synthesis of nucleic acids. Because the main manufacturer has stopped making the equipment we need, we believe that by the end of this year there will no longer be a standard-scale synthesizer manufactured that is really suitable to make DNA or RNA oligonucleotides for research.

Significant areas of RNA research, in particular, require the synthesis of RNA modified in specific functional groups or by the attachment of fluorophores or other reporters. Similarly, antisense oligonucleotide research requires increasingly sophisticated combinations of modified nucleotides and conjugates. For these studies, a research-level synthesizer is indispensable. There will probably be hundreds of laboratories directly affected in the long term, to say nothing of the knock-on effects in the development of the science overall. RNA chemical biology is currently immensely exciting, and moving extremely fast. Yet without the ability to synthesize these molecules at will, the whole field will be very much the poorer.

In our opinion the world leader in the manufacture of DNA/RNA synthesizers has been Applied Biosystems Inc. (ABI). This company produced the venerable 380B and then the 394 research synthesizers. Some time ago it stopped making the 394 synthesizer, and just continued supplying the (in our opinion) inferior Expedite machine. However, it intends to discontinue even that at the end of this year. Thus, although high-throughput machines suitable for preparation of sequencing and PCR primers will still be available, it appears there will be no manufacture of a synthesizer suitable for small-scale, versatile synthesis of nucleic-acid analogues.

No doubt many labs will continue to nurse their existing instruments for some time to come, but as wear and tear take their toll and parts eventually fall into short supply we are likely to lose our ability to make the molecules on which so much exciting research depends. Although ABI has an excellent record of maintaining its obsolete synthesizers, it can give no long-term guarantees of this. Further, if no suitable machines are to be manufactured in the future, this prevents imaginative new researchers coming into the field, which will ultimately have a stultifying effect.

We hope this letter will alert everyone interested in the chemical biology of

nucleic acids to this very worrying development. We also hope that it might stimulate a manufacturer somewhere to fill this gap. Our hope is that companies that grew big by selling to the research community might remember their origins, and not turn their backs on the people who have come to depend upon them for their research.

David M. J. Lilley*, **Michael Gait†**,
Fritz Eckstein‡

*CRC Nucleic Acid Structure Research Group,
Biochemistry Department, MSI/WTB complex,
University of Dundee, Dow Street, Dundee
DD1 5EH, UK

†MRC Laboratory of Molecular Biology, Hills Road,
Cambridge CB2 2QH, UK

‡Abteilung Chemie, Max-Planck-Institut für
Experimentelle Medizin, Hermann Rein Strasse 3,
D-37077 Göttingen, Germany

Outbreak needs lab tests and clinical diagnosis

Sir — Your welcome News report on 12 April (*Nature* **410**, 727; 2001), that the UK foot-and-mouth disease epizootic may be slowing, referred to an article two weeks earlier (*Nature* **410**, 501; 2001). This reported that three independent groups of researchers had told the UK government that rapid slaughter of infected animals was the best way to slow the outbreak.

Without wishing to detract from the valuable contributions being made by epidemiologists, most veterinary clinicians and pathologists dealing with farm animals were aware of this need even before the outbreak started. The need for rapid diagnosis and speedy implementation of control measures for diseases of this type is part of their basic training.

Early delays in dealing with the current outbreak were significantly influenced by owners of diseased animals not requesting veterinary advice quickly enough, possibly because the clinical signs of foot-and-mouth disease, particularly in sheep, can be difficult even for a responsible and observant owner to detect. Another cause of early delay was the requirement to have animals produced for commercial purposes valued — not an issue when dealing with an epizootic in laboratory mice.

At the beginning of an epizootic it is essential to diagnose the disease correctly and identify the causal agent by laboratory tests, to provide epidemiologists with accurate information for mathematical models. In some cases it can take as long as four days to declare a sample positive for foot-and-mouth disease virus.

Once the aetiology of the disease is known, control procedures can be implemented more quickly by relying on

diagnosis based on clinical signs, and laboratory tests are not needed. The effectiveness of this approach depends on the diagnosis being made by people who, in the case of foot-and-mouth, can distinguish between the lesions of this disease and those of the other enzootic diseases prevalent in the United Kingdom such as mucosal disease, bovine papular stomatitis, contagious pustular dermatitis, foot rot and laminitis.

Hugh M. Pirie

North East Corner, Buchanan Castle Estate,
Drymen, Glasgow G63 0HX, UK

Fossil hunters in dispute over Ethiopian sites

Sir — Your News story “Restrictions delay fossil hunts in Ethiopia” (*Nature* **410**, 728; 2001) omitted to report some of the things I told your reporter during our conversation at the American Association of Physical Anthropology meeting in March.

First, the Ethiopian authorities granted me a palaeoanthropological permit last year for the Gadamaitu region, which includes Galili. My team did not displace Yohannes Haile Selassie from this site: his permit was for the Mulu Basin area, which does not overlap with the Galili area. I offered him both cooperation and the opportunity to continue his work with my team in our permit area, but he refused both offers.

Second, when I applied in March 2000 for the second time for a permit for exactly the same site, the Centre for Research and Conservation of Cultural Heritage in Addis Ababa gave me a permit for the Gadamaitu/Galili region for the next three years. If Yohannes Haile Selassie had already held a valid permit for this area, I would not have obtained mine from the issuing authorities.

Third, Ethiopia is a proud, independent country and its authorities are fully capable of writing rules and regulations protecting its cultural heritage. These rules should be respected by foreign scientists; to ask, as your reporter did during our interview, if I had something to do with formulating these regulations is an insult both to me and to Ethiopia.

Ethiopia is working hard at improving democracy and is willing to open its doors to the international scientific community through these new regulations, which I personally fully support.

Horst Seidler

Institute for Anthropology, Althanstr. 14,
1090 Wien, Austria

We stand by the facts as reported in our story — Editor, *Nature*.

BOX 1: Glossary of terminology

Working groups. Writing teams of 100–200 lead authors, dozens of review editors and hundreds of peer reviewers from scores of countries who, as the Intergovernmental Panel on Climate Change (IPCC), produce assessments of the literature for about 100 governments. **Working group 1** deals primarily with the science of climate change, evaluating projections of alternative future climate changes and events. **Working group 2** assesses the potential impacts of such projections, prospects for adaptability and potential vulnerabilities. **Working group 3** assesses mitigation and other policy options for dealing with climate changes. There have been three cycles of assessment reports since 1990.

Special reports. Focused assessments of topics that governments decide need additional treatment. The special report on emission scenarios (SRES) is discussed in this article, but others have been prepared, for example on aviation impacts and carbon sinks.

Storylines. A framework from the SRES explicitly to recognize different demographic, social, economic, technological and environmental developments. Each scenario represents a specific quantitative interpretation of one of four storylines. All the scenarios based on the same storyline constitute a scenario 'family'.

Scenarios. According to the SRES: "Scenarios are alternative images of how the future might unfold and are an appropriate tool with which to analyse how driving forces may influence future emission outcomes and to assess the associated uncertainties."

General circulation models (GCMs). The most comprehensive three-dimensional, time-evolving simulation models available. They often couple submodels of the atmosphere to oceans, to ecological systems and to ice systems. Omission of explicit treatment of processes on a smaller scale than the 'grain size' of the models creates uncertainties in their outputs, especially for regionally specific projections.

Forcing. Any scenario of greenhouse-gas emissions or aerosols implies that the radiant energy transferred between the Earth's surface and space by the atmosphere can be modified. So the scenarios are fed into biogeochemical cycle models to translate the emissions into concentrations of radiatively active constituents, which, in turn, 'force' the climate to change by perturbing the heat balance of the Earth-atmosphere system. Working group 1 has translated SRES emissions scenarios into radiative forcing to drive climate models that in turn produce projections used by working group 2 to assess potential impacts.

What is 'dangerous' climate change?

To combat global warming, we must first assess just how likely it is to occur.

Stephen H. Schneider

In the third assessment report of the Intergovernmental Panel on Climate Change (IPCC), the climate modellers of working group 1 (see Box 1 for glossary) dramatically revised upwards the top-range limit of their predictions of global warming from the previous value of 1–3.5 °C to 1.4–5.8 °C between now and 2100 (refs 1, 2). This sweeping revision depends on two factors³ that were not the handiwork of the modellers: smaller projected emissions of climate-cooling aerosols; and a few predictions containing particularly large CO₂ emissions.

This big increase in the projection of severe climate change is certainly plausible, but an unanswered question lingers: "How likely is it that the world will get 6 °C hotter by 2100?" The answer depends on the likelihood of the assumptions underlying the projections, so the components of this projection deserve careful analysis.

The most famous 'scenarios' (see Box 1) used by working group 1 in the IPCC's second assessment report¹ were six 'business as usual' projections, and one central scenario formed the basis of several calculations used by working group 2 (ref. 4). Because there is a lag of a few years between emissions predictions, climate models' responses, and analysis of possible impacts, the IPCC decided to

prepare a special report on emissions scenarios (SRES) to produce a family of updated projections in time for the writing and review cycle for the third assessment in 1999/2000 (see Box 2, overleaf). For this reason, preliminary scenarios were released, and several climate-modelling groups produced general circulation model (GCM) runs driven by six 'illustrative scenarios' of the 40 emissions projections in the special report. Unfortunately, the time lag persisted, and the impacts-assessment literature⁵ still did not have enough time to incorporate the updated climate runs for the latest round of the IPCC.

The likelihood controversy

I attended a preliminary meeting of the SRES group, led by Nebojsa Nakicenovic, at the International Institute for Applied Systems Analysis near Vienna. I was impressed by the broad representation of the group: academic scientists, environmental organizations, industrial scientists, engineers, economists and systems analysts. Their task was to imagine plausible alternative future societies and technologies that would determine the emissions profiles and drive climate-change scenarios.

The approach — suggested by industrial participants — was to create 'storylines' about future worlds from which population, affluence and technology drivers could be

HUGH TURVEY/SPL

We are facing the worrying prospect of dozens of users selecting arbitrary scenarios and climate sensitivities.

► inferred (see Box 2 for details). It was an impressive exercise, and gave rise to radically different families of emission profiles up to 2100 — from below current CO₂ emissions to five times current emissions. Because of the divergent views of participants about the likelihood of each storyline, the final report offered no assessment of the relative likelihoods, in an attempt to avoid endless disputes.

While acknowledging the logic of avoiding fruitless debate, I strongly argued at the time that policy analysts needed probability estimates to assess the seriousness of the implied impacts; otherwise they would be left to work out the implicit probability assignments for themselves (see ref. 6 for fuller discussion). I urged the expert group to provide a subjective probability assessment for less expert users, but I was not persuasive enough, and the SRES authors expressed “no preference” for each scenario (page 46, ref. 3).

The confusion I feared is already surfacing. The most typical assumption is a uniform probability distribution across storylines (scenarios). This might seem to imply a uniform probability distribution in the outcome that really matters to policymakers: the likelihood of a particular temperature rise by 2100. But this inference would be incorrect, because uncertainties compound through a series of modelling steps. Uncertainties in emissions scenarios feed into uncertainties in carbon-cycle modelling, which feed into uncertainties in climate modelling, which drive an even larger range of uncertain climate impacts. This ‘cascade of uncertainties’⁷ is compounded by the very wide range of emissions offered by the SRES authors.

Working group 1 transferred this broad range of emission projections into radiative forcings of the climate system (see Box 1 and Fig. 18 in the technical summary of ref. 2), producing a wide range of temperature projections by means of a simple model tuned to seven general circulation models (out of a possible 18 cited by working group 1 — Table 9.1 of ref. 2). These seven models represent a range of equilibrium climate sensitivities from 1.7 to 4.2 °C warming for a doubling of CO₂. The result of combining the climate-sensitivity estimates from the seven models with the six illustrative scenarios from the

special report (Box 2) is the dramatic revision upwards in the IPCC’s third assessment which has attracted so much attention, as global warming of more than 3.5 °C would have severe effects⁵.

If all scenarios and sensitivity projections are given equal likelihood, we might intuitively expect that any combination from these two sets would also be uniformly likely, so 1.4 °C would be as probable as 5.8 °C warming. But this is not the case. Instead, the resultant combined probability distribution for global surface warming in 2100 would have a peak at the centre — just like a typical bell curve⁸. If all 18 GCM sensitivities are combined with all six illustrative scenarios (108 possible combinations), then indeed a peaked bell curve results (see Fig. 1 for an example). As a limiting case and for illustration, if only the outlier emissions scenarios are combined with the 18 climate sensitivities (36 possible combinations), then the probability distribution for 2100 warming would not be as peaked as a bell curve, but would have a much flatter distribution.

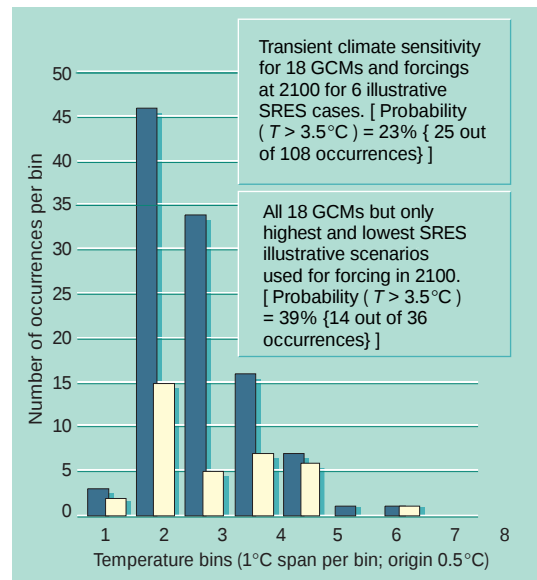
For illustrative purposes, I have chosen arbitrarily a temperature increase of 3.5 °C as a ‘threshold’ beyond which many believe substantial climate damage would occur. In Fig. 1, the blue bars show that the resultant bell curve for all 18 GCMs and all six illustrative scenarios implies that 23% of the 108 values

obtained for a possible temperature increase in 2100 are greater than 3.5 °C. For the truncated case in which only the highest and lowest emissions scenarios are used (yellow bars), 39% of these 36 possibilities are for a warming of 3.5 °C or greater. With the relative lack of middle-value scenarios, the probability distribution is much flatter than before.

Clearly, a policymaker concerned with “avoiding dangerous anthropogenic interference in the climate system”⁹ would propose stronger policies and measures if there was a 39% chance of exceeding the 3.5 °C warming ‘threshold’ than if the figure was 23%. But what do these figures actually represent? Unless probabilities are assigned to individual scenarios and GCM climate sensitivities (for example, as in ref. 10), their joint distribution (in this example the likelihood of some temperature rise in 2100) will depend on the particular selection of scenarios and models, as Fig. 1 clearly demonstrates. To assess the likelihood of future temperature increase, we can either estimate more consistently the subjective likelihood for each scenario and climate sensitivity, or estimate the joint distribution (2100 temperatures) explicitly. Early attempts at this estimation have been made (T. Wigley and S. Raper, personal communication; M. Webster *et al.*, see <http://mit.edu/globalchange/www/reports.html>).

Figure 1 ‘Frequency’ of ‘severe’ climate impacts in 2100. The difference between the yellow and blue bars shows how selection of scenarios can lead to arbitrary results. The histograms are the number of occurrences of temperature increases around 2100 in seven bins, each 1 °C wide and beginning at 0.5 °C (for example bin 3 represents the number of occurrences of temperature increases (*T*) between 2.5 °C and 3.5 °C at 2100). The blue bars represent the distribution that results from the product of all 18 GCM transient climate sensitivities (in °C warming for a transient climate forcing at the time of an equivalent doubling of CO₂; from Table 9.1 of ref. 2); and all six radiative forcings at the year 2100 from the six illustrative emissions scenarios (from Fig. 18 of the technical summary of ref. 2).

Of the 108 numbers represented by the joint inclusion of 18 climate sensitivities and six forcings, 25 occur in bins representing temperature increases greater than a threshold value of 3.5 °C — a rise that many would consider could cause extensive and significant climate damage. The yellow bars represent a case in which all 18 GCM sensitivities are used, but only two emission scenario forcings (the highest, A1FI, and the lowest, B1). Fourteen out of 36 outcomes are greater than the threshold of 3.5 °C warming in 2100, representing a much larger likelihood (39%) of severe climate damage than for the other case. The likelihood of crossing some threshold is thus sensitive to the particular selection of scenarios and climate sensitivities used. Arbitrary selections will produce distributions that could easily be misinterpreted as containing subjective probabilistic analysis when they do not — until judgements are formally made about the likelihood of each scenario or sensitivity. For this reason the word ‘frequency’ appears above in quotation marks. It is not a justifiable probability distribution as the subcomponents are chosen without a “traceable account”⁶ of the logic of the selection process.





Box 2: Summary of special report's findings

The most compelling conclusions of the special report on emissions scenarios are:

(1) Alternative combinations of main scenario driving forces can lead to similar levels of greenhouse-gas emissions by the end of this century. Scenarios with different underlying assumptions can result in very similar climate changes.

(2) Technology is at least as important a driving force of greenhouse-gas emissions as are population and economic development across the set of 40 scenarios.

The 40 scenarios, consisting of four main 'families' and various subgroups, each have a distinct narrative storyline containing different assumptions on population, economic and technological growth; orientation towards a global economy and community; and attitudes towards economic, social and environmental goals.

The scenarios cover the full range of greenhouse-gas and SO₂ emissions that SRES authors could imagine as plausible.

One prominent scenario group is the A1 series. This is distinguished by its technological emphasis on coal (A1C), oil and gas (A1G), non-fossil energy sources (A1T) or a balance across all sources (A1B). In the working group 1 summary for policymakers, A1C and A1G groups are combined into one fossil-intensive group, A1FI. All scenario families are said to be 'equally sound'.

The scenarios are also grouped into four categories of cumulative CO₂ emissions, which indicate that scenarios with different driving forces can lead to similar cumulative emissions, and those with similar driving forces can branch out into different categories of cumulative emissions.

Four scenarios are designated as 'markers'. Together with two scenarios from the A1 family, the six 'illustrative' scenarios form the backbone of the range reported by working group 1.

The report's authors recommend that any evaluation should include at least the six illustrative scenarios.

All scenarios describe futures that are generally more affluent than today. Many envisage a more rapid convergence in per capita income ratios in the world than do the 'business as usual' scenarios, despite having divergent greenhouse gas and SO₂ emissions (compare A1 and B1 scenario families).

Many greenhouse-gas emissions described in the special report are lower than the central 'business as usual' level in the second assessment report, especially towards the end of the twenty-first century. Emissions of SO₂, which have a cooling effect on the atmosphere, are significantly lower than in the second report.

Emissions, impacts and the implications of policy responses all depend on how society is structured over time.

Social conditions and emissions

One of the most important points of the SRES storylines approach is that the socio-economic conditions driving emissions would also help to form the adaptive and mitigative capacity of various countries or regions¹¹. Emissions, impacts and the implications of various policy responses are all mutually dependent on how society is structured over time. This insight is one of the most positive developments of the approach, and will occupy integrated assessments of climate change for the foreseeable future. It is essential that the storyline approach should continue and expand, but, as I have shown here, future work should attempt to deal more forthrightly with the difficult issue of the relative likelihood of each scenario, and to provide more guidance as to the independence of each storyline.

What next?

The special report leadership was not wrong, of course, about how difficult it would be to assign subjective probabilities to radically different visions of the future. But in the probability vacuum that followed its assertion that all scenarios were "equally sound", we are facing the even more worrying prospect of dozens of users selecting arbitrary scenarios and climate sensitivities to construct frequency charts that, like the histogram of Fig. 1, are built on implicit assumptions. In the risk-management dilemma that constitutes climate-change policymaking, I would definitely put more trust in the probability estimates of the SRES team — however subjective — than those of the myriad special interests that have been encouraged to make their own selection.

Meanwhile, as we wait for the IPCC to decide whether to reassemble the team for

this controversial labour, climate policy-makers will have to be vigilant, asking all advisers to justify the threshold they choose for predicting 'dangerous' climate change, as well as to provide a "traceable account"⁶ of how they selected their emissions scenarios and model sensitivities, as these jointly determine the probability of future risks. ■
Stephen H. Schneider is in the Department of Biological Sciences, Stanford University, Stanford, California 94305, USA.

1. Intergovernmental Panel on Climate Change *Climate Change 1995*. Contribution of working group I to the second assessment report of IPCC (eds Houghton, J. T. et al.) (Cambridge Univ. Press, 1996).
2. Intergovernmental Panel on Climate Change *Third Assessment Report of Working Group I: The Science of Climate Change* (Cambridge Univ. Press, in the press).
3. Nakicenovic, N. & Swart, R. *Special Report of the Intergovernmental Panel on Climate Change on Emissions Scenarios*. (Cambridge Univ. Press, 2000). Summary for Policymakers available online at <http://www.ipcc.ch>
4. Intergovernmental Panel on Climate Change *Climate Change 1995*. Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change (eds Watson, R. T., Zinyowera, M. C. & Moss, R. H.) (Cambridge Univ. Press, 1996).
5. Intergovernmental Panel on Climate Change *Third Assessment Report of Working Group II* (Cambridge Univ. Press, in the press).
6. Moss, R. H. & Schneider, S. H. in *Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC* (eds Pachauri, R., Taniguchi, T. & Tanaka, K.) 33–51 (World Meteorological Organization, Geneva).
7. Schneider, S. H. in *Social Science Research and Climate Change: An Interdisciplinary Appraisal* (eds Chen, R. S., Boulding, E. & Schneider, S. H.) 9–15 (Reidel, Boston, 1983).
8. Jones, R. N. *Climatic Change* **45**, 403–419 (2000)
9. United Nations Framework Convention on Climate Change, Rio de Janeiro (United Nations, 1992). Text available on the UNFCCC Secretariat web site: <http://www.unfccc.int>
10. Morgan, M. G. & Keith, D. W. *Env. Sci. Tech.* **29**, 468A–476A (1995).
11. Intergovernmental Panel on Climate Change *Third Assessment Report of Working Group III* (Cambridge Univ. Press, in the press).

Acknowledgements

I thank C. Azar, K. Kuntz-Duriseti, M. MacCracken, R. Moss, S. Raper and T. Wigley for comments on earlier drafts.

Volcanic irony

Hard lessons from the dark side of volcanology.

Surviving Galeras

by Stanley Williams & Fen Montaigne
Houghton Mifflin: 2001. 270 pp. \$25

No Apparent Danger: The True Story of Volcanic Disaster at Galeras and Nevado del Ruiz

by Victoria Bruce
HarperCollins: 2001. 239 pp. \$26

William I. Rose

At an international workshop at Galeras volcano in Colombia in 1993, nine people, including six volcanologists, died in a small summit eruption. This disaster is the subject of both these books. One is co-authored by Stanley Williams, a senior volcanologist who was a co-organizer of the workshop and who was severely injured in the accident; the other is an outsider's look by Victoria Bruce, also a geologist, who paints a different picture of the same events. Both books are well written portrayals of scientists doing exciting work. They can be absorbed quickly — I read each of them in a night. All volcanologists and many other scientists should read them, because they raise a number of vital questions.

What is the role of gallantry (bravery?) in volcanology (and in science in general)? What safety issues and rules of ethics ought to be discussed and advanced to try to stem the rising numbers of volcanological deaths? How are the newest theories tested by scientists who compete for recognition? How can the knowledge and technology for mitigating natural hazards best be transferred to less-developed countries?

Volcanology apparently attracts gallant daredevils. Stanley Williams explains, "Most people flee from erupting volcanoes. We head straight for them. ... My colleagues and I don't harbor a death wish. ... the goal, which has taken me to more than 100 volcanoes in two dozen countries, is a worthy one: to improve our ability to forecast eruptions."

Williams also makes a big point about "my kind of scientist", an assertion that excludes those who are the other kind. Williams' kind are the brave and courageous, who want most to be inside the crater. "All the volcanologists I admire, whether they've died in eruptions or lived to old age, share a passion for working on volcanoes. ... we're also hooked on the thrill of climbing into the crater, of confronting so monumental a force. No place on earth leaves me feeling as alive as a volcano does." Williams' kind doesn't appear to include seismologists: "despite the progress we've made in taking a mountain's measure using seis-



Dangerous job: a volcanologist takes gas samples from a fumarole on Vulcano Island, Italy.

mometers and other remote sensing devices, the best way to understand a volcano is to climb it."

These statements express what one could call the dark side of volcanology. They suggest that bravado by itself is admirable. That work in the danger zone is more worthy. That we volcanologists mostly do our work because we get a charge out of seeing red rocks fly over our bare heads.

Although I too have spent many years working in active craters, I think that volcanology is not well served by machismo. The lesson that we should learn from the rapid increase in volcanological deaths is that the pain and suffering resulting from these accidents is excruciating. Both books agree on this point and document it well. Having worked with dozens of graduate students, I have found that it is far too common for people to be attracted to the field because they want to be near the molten

magma — "it's so cool!" Perhaps after reading these books we can roundly reject the professional gallantry that seems to raise risk-taking to the level of heroism. The truly gallant people in this story were Marta Calvache and Patricia Mothes, two women geologists who went into the danger zone on the fateful day only to rescue their friends and colleagues.

In these days of safety regulations for everything, it will surprise many that volcanologists would enter active craters without the safety gear that is *de rigueur* in the workplace (hard hats, flame-proof clothing, gas masks, goggles). Not only did this happen at Galeras, but afterwards Williams stated: "Even if I had forced everyone to put on protective clothing, I am confident that not a single life would have been saved."

This statement, even if true, overlooks much. I doubt that anyone in that crater that day would have regretted having a

ROGER RESSMEYER/CORBIS

hard hat. Bruce describes the ironic observations of the Pastusos (the residents of Pasto, the city near Galeras). They had lived for decades near the volcano without many fatal accidents, yet saw how eagerly volcanologists went to the only place that was really dangerous. Did these researchers really know much? Could they be trusted to tell Pastusos how to live with the volcano, or when to evacuate their homes? Tragically, three Pastusos actually followed the scientists into the crater, encouraged by their nonchalance. The tourists all died.

It is ironic that the educated experts who were going to develop a safety plan for the public paid so little attention to their own safety. Safety measures are at least partly for show — to make people aware of danger. Because often a volcano erupts less than once in a human lifetime, building public awareness of danger is especially crucial in this case. If scientists had been seen in protective gear by those watching near the crater or on television, it would have helped to show that the experts were aware of possible danger. I doubt that any international hazards workshop will ever again forget this.

Williams' comments on seismology also have a tragic irony, as long-period seismic signals were warning of an impending eruption, if anyone had acknowledged the fact. It was realized very soon after the eruption, and both books discuss this at length. Long-period earthquakes are associated with rising magma and degassing, and occur hours or days before an eruption. They were scientific 'hot buttons' at the time of Galeras.

Bruce points out that Williams was aware of this. He had previously mentioned the need to correlate seismic and gas data in a proposal written after a meeting in Pasto, where Bernard Chouet of the US Geological Survey explained to Williams and others that long-period events could be useful for inferring gas-related phenomena and predicting eruption. Williams claims that he realized only after the 1993 eruption that long-period events might presage an eruption. But seismic signals of this type had preceded a Galeras eruption in July 1992. Bruce notes that John Stix had written an abstract to be presented at the January 1993 workshop that recognized the analogy between the July 1992 and January 1993 situations.

So why were the long-period signals not noted and recognized as a warning before the January 1993 crater visit? It seems that the scientists knew of the new ideas — but didn't really trust them or use them. They would cite them in their proposals and abstracts, but when it came to deciding whether to go to the crater, they didn't think them important. Scientists from the US Geological Survey, who might have forcefully pushed Chouet's ideas, missed the Galeras

meeting because they couldn't get travel clearance. Was their absence a factor in the disaster? Was it bravado, scientific rivalry, or just ignorance, as Williams asserts, that caused the warning to be overlooked? If it was rivalry, how can we minimize this in similar circumstances?

The existence of two books about the same events makes them much more valuable, because they each present the events of the Galeras workshop and the fatal eruption from a very different perspective and with different emphasis. Williams describes the lure and lore of volcanology, and I suggest you read him first. He portrays in some detail the people who died; these were Williams' kind of volcanologists, and all remarkable. Igor Menyailov is Williams' ultimate hero; Bruce, on the other hand, criticizes Menyailov as reckless.

In her book, Bruce forcefully points out Williams' mistakes (although, in fairness, Williams recognizes some of them too), using careful interviewing and library research. She focuses on Colombia's volcanologists and the development of an infrastructure to deal with volcanic hazards there. She also considers the Nevado del Ruiz disaster of 1985, in which more than 23,000 people died. This was a most spectacular failure of efforts to mitigate volcanic hazards — a much more tragic event than Galeras.

For those of us trying to help with technology transfer and infrastructure building for natural hazard mitigation in less-developed countries, Bruce's detailed examination of this disaster is praiseworthy. Her thoughtful discussion makes it clear that Colombia lost by far the most from both the Ruiz and Galeras incidents. The increasing emphasis on foreign work in natural hazards areas makes intercultural communications a vital skill. Volcanologists have much to learn here.

This review may seem highly critical of Williams, but that is largely bad luck (his). He was unlucky in that he was the senior scientist, who had convened the workshop, and therefore was held responsible for the disaster. Many of us could easily have been in the same shoes. All who have been in similar situations have made mistakes, which, in hindsight, put people in danger that we and they didn't fully appreciate. We never had to own up to these mistakes because nothing bad happened on our watch. My hope is that volcanologists (and other scientists) read these books and recognize the mistakes that might haunt their profession. As Williams says, "people in high risk professions are confident they can beat the odds". ■

William I. Rose is in the Department of Geological Engineering and Sciences, Michigan Technological University, 1400 Townsend Drive, Houghton, Michigan 49931-1295, USA.

Too faint a flavour?

Flavour and Fragrance Chemistry
edited by Virginia Labzotti
& Orazio Taglialetela-Scafati
Kluwer Academic: 2000. 256 pp.
£74.50, \$117.50

Peter Waterman

The chemistry underlying our perceptions of flavour and fragrance is both fascinating and complex. Although fragrance phenomena are predominantly associated with relatively small, volatile molecules, usually occurring in complex mixtures, the structural range and size of compounds associated with flavour are far greater.

Given that the book under review is volume 46 of the *Proceedings of the Phytochemical Society of Europe*, a series that has produced many notable publications on natural products, I was anticipating a 'good read' to bring me up to date on these topics. Unfortunately, I was more than a little disappointed.

Why am I being so negative? It is not primarily due to the quality of the individual contributions, although they are variable. But with 23 papers and only 250 pages of text, the average length of each contribution is less than 11 pages and the individual papers are, for the most part, simply reports of the work (generally previously published) of the authors. This means we are offered tantalizing fragments on individual subjects, but rarely placed in any review context. The editors have attempted to arrange the chapters according to subject area, but I did not find that their efforts helped to rationalize the disparate topics. It also has to be said that some of the topics have only a peripheral bearing on the subject (for example, alkaloids from Amaryllidaceae, cell cultures of *Hypericum perforatum* for high-value metabolites, antioxidants).

I cannot reconcile the content or the depth of the papers in this volume with my expectations for the flagship series of a prestigious society. When I pick up a volume of the *Proceedings of the Phytochemical Society of Europe* I expect to find erudite and comprehensive reviews of the title subject. This collection of worthy but limited contributions would be better suited to a 'special issue' of an appropriate journal or a cheaper, soft-covered, conference report.

I would not recommend this book to a hard-pressed librarian trying to make ends meet. It is, quite simply, not worth the price. ■

Peter Waterman is at the Centre for Phytochemistry, Southern Cross University, PO Box 157, Lismore, New South Wales 2480, Australia.



Scrapped: left, UN weapons inspectors in 1996 destroy the Al-Hakam building in Iraq, which could have been used to produce biological weapons. Above, a decommissioned Russian nuclear submarine in the Arctic base of Severomorsk in 1998.

AP

Who guards the guardians?

The Verification Yearbook 2000
edited by Trevor Findlay
VERTIC: 2000. 288 pp. £30

Nicholas Zarimpas, Jean Pascal
Zanders and Zdzislaw Lachowski

In a period of apparent stagnation in arms control, the *VERTIC Verification Yearbook 2000* is very welcome. This special millennium edition surveys the evolution of verification over a wide range of fields — from arms control and disarmament to less traditional areas such as the monitoring of environmental treaties and agreements in post-cold-war conflicts.

The control of nuclear weapons is probably the aspect of arms control that springs most readily to mind. International nuclear safeguards are aimed at ensuring compliance with the terms and objectives of the 1968 Treaty on the Non-Proliferation of Nuclear Weapons (NPT). Non-nuclear weapon states accept safeguards such as international inspections and monitoring to show that they are not using their civilian nuclear facilities for military purposes.

David Fischer's chapter "Nuclear safeguards: evolution and future" provides an eloquent historical overview of the application of safeguards. He focuses on the challenges that confronted the International Atomic Energy Agency verification system in the 1990s. Iraq, in violation of its obligations under the NPT, had pursued an extensive clandestine nuclear weapons programme. The nuclear ambitions of North Korea and its non-compliance with the safeguards laid down by the International Atomic Energy Agency (IAEA) posed

yet another problem. Such problems were not easy to address, principally because of limited resources, but also owing to a growing list of responsibilities for the agency, including a potential role in the nuclear disarmament process. Nevertheless, the IAEA and its member states reacted swiftly. Their efforts to strengthen the global nuclear verification regime are central in Fischer's analysis.

Annette Schaper, in "Verifying nuclear arms control and disarmament", expertly guides the reader through this highly esoteric and technically complex subject. Her contribution first focuses on the destruction of delivery vehicles and launch systems. After reviewing experience gained with verifying a range of bilateral US–Russian treaties, Schaper discusses transparency and controls in warhead dismantlement, a logical and indispensable next step in nuclear arms control, as well as the daunting problems surrounding verification of a world free of nuclear weapons. Her contribution concludes optimistically: "it is wise to design verification as if a nuclear weapon-free world is the objective, even though a decision as to whether it should become reality may be delayed."

Chemical and biological weapons are the other main categories of non-conventional weapons with the potential to cause mass casualties. Robert Mathews, scientific adviser to the Australian delegation, provides a valuable assessment of the development of the Chemical Weapons Convention (CWC) in his chapter "Chemical disarmament: advent and performance of the OPCW". He captures the inherent tension between the disarmament and cooperation provisions of the treaty and provides examples of how these tensions have played out.

There has been frustration about the use of the treaty by some states for political

purposes, which has complicated issues such as the amendment of transfer procedures for certain scheduled chemicals with a clearly medical or diagnostic end use. At the same time, however, there have been encouraging increases in the cooperation shown on issues such as compliance and conflicting information submitted by 'states parties'.

Some of the other unresolved political and technical issues that Mathews deals with are now acute, such as the timely destruction of chemical weapons in Russia. He rightly concludes that, despite many political compromises and teething problems, the CWC has set up a verification regime that functions as intended, and is able to provide confidence that states parties are complying with the treaty provisions.

The Biological and Toxin Weapons Convention (BTWC) was the first true disarmament treaty obliging states parties to eliminate an entire category of weapons, but it lacks even the most basic verification regime. Now that the CWC has come into force, and in the light of concerns about Russia's compliance and revelations about Iraq's biological weapons programme, the BTWC is viewed as a weak treaty.

Verification and transparency measures are urgently needed. Nicholas Sims sketches the evolution of the BTWC regime via the process of quinquennial review conferences: states parties have gradually accepted functional substitutes, such as confidence-building measures and politically binding obligations to report on treaty-relevant activities, for verification. Although these keep the treaty alive, the need for true verification remains imperative.

The chapter "Verification of conventional arms control" by Pál Dunay is a valuable review of this aspect of arms control in Europe. Confining his article to structural

arms control (that is, weapons limitation and reduction), Dunay analyses the accomplishments of various forums over the past half-century: control over West Germany, the Mutual and Balanced Force Reduction talks, the verification regime for the Treaty on Conventional Armed Forces in Europe and the regional approach in the Balkans.

Today, arms control arrangements are increasingly focused on regional contexts and have less to do with weapons limitation than with confidence and security-building measures, transparency (such as clarification of the military information provided, information on the methodology of defence planning, high-level military doctrine seminars), risk reduction and stabilization. This may mean that the stringency of verification will gradually give way to broader, less binding regimes within the cooperative security system in Europe.

We cannot yet tell how history will judge the efforts to eliminate Iraq's arsenals of non-conventional weapons following its eviction from Kuwait in 1991. The UN Security Council ordered the destruction of Iraq's chemical and biological weapons as part of the cease-fire resolution 687 (1991) and to this end created a Special Commission — UNSCOM. Stephen Black, who was attached to UNSCOM as a historian, neatly describes the establishment of the verification regime, its expansion and successes, and the political difficulties and compromises that eventually led to the dissolution of UNSCOM in 1999 and its replacement with the still inactive UN Monitoring, Verification and Inspection Commission (UNMOVIC).

Although Iraq represents a case of coercive disarmament in an extremely hostile environment, several valuable lessons can be learned. Most importantly, as the author notes, parties to arms control and disarmament treaties must understand the need for continued active involvement in the treaties. Indeed, the UNSCOM saga also tells us that the UN Security Council is unable or unwilling to uphold its own resolutions despite blatant violations by Iraq, and that certain permanent members easily succumb to short-term self-interest. As the UN Security Council is the ultimate arbiter for material breaches of disarmament treaties, this is a serious cause for concern.

The *Yearbook* represents a serious and concerted effort to introduce the reader to the evolving and multifaceted world of verification. We recommend it to practitioners, academics, and the less technical reader alike. ■

Nicholas Zarimpas, Jean Pascal Zanders and Zdzislaw Lachowski are at the Stockholm International Peace Research Institute (SIPRI), Signalistgatan 9, S-169 70 Solna, Sweden.

Science in culture

Dramatizing science

Oxygen, a play by Carl Djerassi and Roald Hoffmann
Phillip Ball

Conveying the language of any specialized profession in a dramatic context is never easy. But can that alone explain why it is so hard to portray scientists realistically on the stage and screen?

On the printed page we can somehow accept much more: Alice in Jeanette Winterson's *Gut Symmetries* or Thelma Darke in Ian McEwan's *The Child in Time*, both theoretical physicists, don't sound gauche as they hold forth on relativity. But put scientific characters on stage and their words become, in the ears of professional scientists, stilted and false.

Only the finest playwrights seem able to avoid this trap — and then by sacrificing all but the appearance. Tom Stoppard's *Valentine Coverly in Arcadia* gives us an elegant and convincing exposition on chaos, but, like Jeff Goldblum in the film *Jurassic Park*, you know he has never really spent hours plotting Poincaré maps. Bertolt Brecht's Galileo works so well as a character because he seldom has to speak as a scientist.

When playwrights do try to show us scientists speaking to one another, we never hear the rhythms of lab speech, but instead an awkward counterfeit — even when the writer is as well informed as Stephen Poliakov, whose *Blinded by the Sun* was inspired by the cold fusion episode and who is the brother of a chemist.

Oxygen, a play by Carl Djerassi and Roald Hoffmann premiered in San Diego last month at the meeting of the American Chemical Society, is interesting because it reveals the professional scientist's attempt to bring off this dramatic sleight-of-hand. Here, at least, we know that the chemistry will be impeccable:

Ulf Svanholm: You remember the Stanford group's paper on new catalysts for oxygenated polymers?

Bengt Hjalmarsson: Didn't you have some similar catalysts up your sleeve?

Ulf Svanholm: Identical. Except that the American paper came out several months earlier.

Let no one underestimate the preparation needed for an actor to say "oxygenated polymers" and sound as though he not only knows what they are but has been working on them for the past five years. Playwrights take a great risk when they use language like this — not Stoppard-style explanations to outsiders, but chat between peers.

But that is part of the game for Djerassi, who confesses to the aim of using theatre to 'smuggle' some science into people's field of view. His concept of 'science in theatre' is explicitly pedagogical — the audience will leave not only entertained, but also better informed.

At face value, this sounds akin to Brecht's didactic theatre. But Brecht's motives were moral and political: he set out to persuade us of a certain stance or argument, not to deepen our



Whodunnit?

historical knowledge of, say, the Thirty Years' War. *The Life of Galileo* does not seek to teach us heliocentric astronomy, but to question the relationship between person and state, between enquiry and ideology.

This is why Djerassi's 'science in theatre' has one foot firmly in science education. That is a valuable role, and *Oxygen* should be an effective emissary. But we should not mistake the play for something else. It entertains, it informs, it has the clever device of marshalling the evidence (the question is: who discovered oxygen?) and letting the audience reach their own conclusions. But at times, for example when discussing the phlogiston theory, the characters are clearly providing explanations more for the audience's sake than for one another's. And in terms of probing, say, the counterpoint of truth and ambition, the play reaches no further than *Blinded by the Sun* (which is not far).

The story rests on the premise that the Nobel Committee of 2001 are awarding 'retro-Nobels' for great discoveries made before the prizes began. In chemistry, the first is to go to oxygen's discoverer — but should that be Carl Wilhelm Scheele, Joseph Priestley (both of whom made it but were avowed phlogistonists) or Antoine Lavoisier, whose experiments came last but who reached the correct interpretation? The action jumps back and forth between the chemistry committee's deliberations and a meeting between the three protagonists in Stockholm in 1777.

Perhaps it is to other scientists that the play will be most educative, as it shows how little the squabbles of Priestley, Scheele and Lavoisier matter for chemistry today, and how arbitrary is any final attribution. One hopes that scientific audiences will not respond like the fictional Nobel Committee, grudgingly acknowledging this truth and then, by each advocating his own candidate until the bitter end, proceeding to ignore it. ■

Philip Ball is a consultant editor of *Nature*.

Oxygen is published by Wiley-VCH. It will be performed (in German) at the Stadttheater in Würzburg from 23 September and in London (in English) in November at the Royal Institution.

Newton in Japan

Translators of scientific texts add their own, cultural interpretations.

Scott L. Montgomery

Isaac Newton wrote in Latin and in English. Today his works, and the vocabulary they introduced, are found in every major language of the world. Are all these versions of the 'founder of modern physics', roaming the shelves and classrooms of the planet, essentially the same? Is the terminology of 'force', 'velocity' and 'acceleration', for all practical purposes, identical in tongues otherwise as different as English and Japanese?

The story of Newton's entry into Japanese is an interesting one. The man responsible for this epochal feat was Shizuki Tadao, a *rangaku-sha* (scholar of Dutch studies), who, between 1798 and 1802, composed a work of three volumes in which Newtonian physics and astronomy formed an essential core. The title of this work is revealing: *Rekishō Shinsho*, "New Writings on Calendrical Phenomena", indicating allegiance to the traditional focus of astronomy in Japan, which had been guided by Chinese neo-Confucian natural philosophy for centuries. But Shizuki's title was also something of a screen, behind which lay a revolutionary set of introductions that would help overturn neo-Confucian science even while incorporating portions of it.

Shizuki composed his work at a crucial moment in the history of Japanese science. From 1630 to 1720, no Western books had been allowed to enter the country, and exchange with Europe was restricted to trading with the Dutch, who continued as the sole supplier of things Western up until the nineteenth century. Only after about 1770 did scientific books become a significant part of this exchange. These were handled, and either partly or wholly translated, by the *rangaku-sha*, who were employees of the government. The Tokugawa shogunate remained suspicious of Western texts, and especially condemned any reference to religious matters or the word "God" (thus making scientific books the safest to trade). But by the 1790s, when Shizuki was active, a few translations of European medical and astronomical works had finally been published and made available, allowing scholars access to new styles of explanation, drawing, and prediction. A burgeoning interest in European science resulted among Japanese intellectuals.

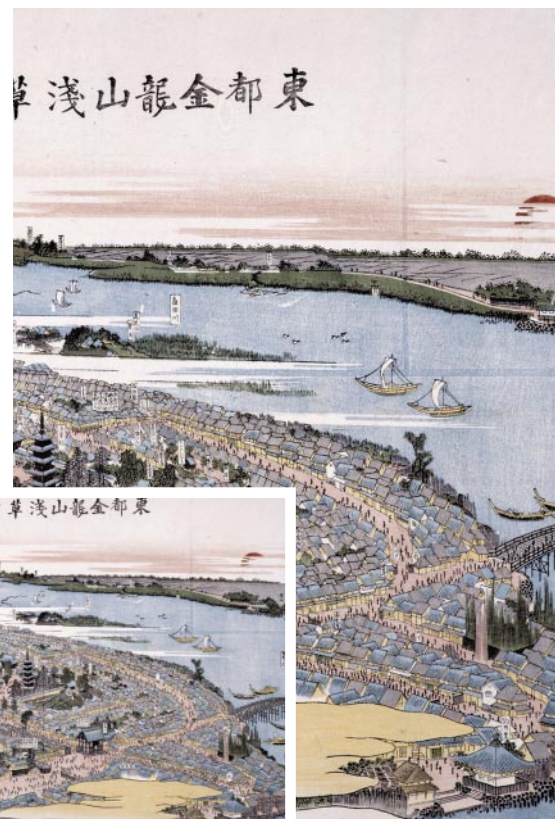
The larger situation was thus uncertain. Shizuki would have been ill-advised to use the real name of his principal source: *Inleiding tot de Waare Natuuren Sterrekunde* —

"Introduction to the True Natural Philosophy and Astronomy". This was itself a Dutch translation from an original popular-science book by the English author John Keill, one of Newton's most ardent promoters. But what did Shizuki do with it?

In Western terms, Shizuki might be called both an 'ancient' and a 'modern'. As an official interpreter and a scholar of the Chinese classics (as nearly all scholars were), he felt the need to interpret the heliocentric view in accord with the ethical prerogatives of neo-Confucianism, to wit: "There exists a governing centre in all things: for an individual, the heart; for a household, the father; for a province, the government; for the whole country, the imperial court; and for the entire universe, the Sun." Thus could a Sun-centred cosmos be made acceptable.

What of his translations of newtonian terms? Here Shizuki showed himself a modern thinker of great ingenuity. For many basic words — force, gravity, velocity, elasticity, attraction — he employed a simple system. This involved using the ideogram for a related action in each case — movement, weight, speed, stretch, pull — and combining it with the ideogram for strength or power (*chikara*). Thus force became 'movement power' (power to create movement); gravity was 'weight power'; velocity 'speed power', and so forth. Each term thus gained a concrete sensibility, a link to everyday experience, and also a degree of self-explanation largely lacking in the English or Latin. For other terms, Shizuki returned to the neo-Confucian influence: vacuum rendered as *shinkū*, 'true emptiness'; corpuscle as *bunshi*, 'child part' (which later became the term for molecule); centripetal and centrifugal, as *kyūshin* and *enshin*, meaning 'seeking the heart' and 'receding from the heart'.

All these terms remain in use today, as part of the basic vocabulary of modern physics. Indeed, later translations of Newton's own books were forced to obey Shizuki's choices, so widely had they been adopted. Moreover, they set patterns that were followed in coining new terminology in the nineteenth and twentieth centuries. Newtonian language thus condenses a moment when loyalties to both China and the West were in motion. To the knowing eye or ear,



Ancient city, new ideas: Tokyo in Shizuki's time.

the daily use of this language today rehearses such a moment.

For most of the eighteenth century and well into the early nineteenth, translation was the dominant work performed by people involved in Japanese science. 'Scientists' we may call them, but they were translators first, and because of this, their influence was profound. The creation of a new vocabulary in science is no mean feat. To a degree, Shizuki Tadao deserves a place beside Lavoisier, Linnaeus and their peers.

How, then, to answer our initial question? Is Newton's terminology the same in all languages? The case presented here suggests that there are, in fact, a host of 'Newtons' roaming the globe today — that newtonian language has been adapted significantly to each new linguistic context, not the other way around. Can this be said to challenge the paternity of this great 'father' of modern physics? Only if we demand that such an image must be absolute, that it must erase all those workers who made newtonian thought a living thing in other languages. For Newton is today much larger and more polymorphic than he was in the past. The Japanese example teaches us that his intellectual parenthood is as much a matter of what the world has done with his legacy as of its original content. ■

Scott L. Montgomery is a geologist and independent scholar. His most recent books include *Science in Translation* (University of Chicago Press, 2000). e-mail: scott.montgomery@prodigy.net

Here be no dragons

David Ruelle

Nature presents us with a bewildering variety of complex systems, from the Earth's atmosphere, to the stock market, to the anatomy of the fly. Their complexity appears in different ways and for different reasons. One cannot hope for a general theory to explain all complex systems, but theoretical ideas may provide insight into some. The idea of biological evolution helps us understand the complexity of living beings; the idea of chaos helps make sense of some complicated, irregular, time evolutions such as the flow of a turbulent fluid.

How can you obtain oscillations with a complicated, irregular, time dependence? One way (but not the most interesting) is to put together many oscillators with different frequencies. Another is to use just a few oscillators (say three), and throw in a nonlinear interaction, resulting in a deterministic time evolution with sensitive dependence on initial condition. What is that? Well, in a deterministic time evolution, if you know the state $x(0)$ of the system at time 0, you can predict its state $x(t)$ at time t with complete precision. But if there is an imprecision $\epsilon(0)$ in the initial condition $x(0)$, the imprecision $\epsilon(t)$ in $x(t)$ may grow exponentially with t , at least for a while. This is sensitive dependence on initial condition — in the case of a pencil one is attempting to balance on its tip, the initial condition determines the side to which it falls. The surprise is that in some systems there is sensitive dependence on initial condition whatever the initial condition. This is called chaos: a chaotic system is unstable everywhere. Obviously, the evolution in time of a chaotic system has limited predictability, and the system undergoes

irregular oscillations (if they were regular they would be predictable).

The mathematical possibility of chaos was understood 100 years ago by Hadamard and Poincaré, but theory and applications were developed more recently. Apart from a remarkable paper by Lorenz in 1963, the flow of papers started in the 1970s. Computer studies then made it clear that chaos is rather a common phenomenon. We now know that hydrodynamic turbulence is chaotic, which certainly puts the subject in a new perspective. We also know that the motion of the Earth's atmosphere has sensitive dependence on initial condition, which sets fundamental limits on weather predictions. (Although we have good knowledge of short-term weather evolution, understanding longer-term climate evolution largely escapes us.) Among the more recent successful applications of chaos theory, the most spectacular relate to the Solar System: the parameters of the orbit of the Earth vary chaotically with a characteristic time of about five million years (short on geological timescales), which has implications for palaeoclimates.

The basic concepts and methods of chaos have become widely accessible, being successfully applied to various interesting, relatively simple, precisely controlled systems. Other attempts, for example applications to the stock market, have not yielded convincing results: here our understanding of the dynamics remains inadequate for a quantitative application of the ideas of chaos.

There are philosophically important consequences of chaos. Poincaré noted that the uncertainty of our predictions justifies a probabilistic description of the world. But we can now go further: we can compute the probabilities of some historical events, such as the chance of rain on Piccadilly Circus or Times Square, given what is known of the state of the Earth's atmosphere 48 hours earlier.

This leads me to digress on the tantalizing topic of historical probabilities. Philosophers and scientists have speculated about the probability of life arising on Earth or elsewhere. Can one construct a probabilistic history of the world — the formation of the Solar System, life on Earth, or human history? In principle one can speak about the conditional probability of a certain event, given what is known about the state of the Universe at some earlier time. Two things are needed for such a probability to be of interest: first, that it can be estimated at all; and second, that the estimate doesn't depend too much on insignificant details of the assumed earlier state of the Universe. The estimate can involve not just chaos theory, but other ideas of smooth dynamics, quantum uncertainty,

Chaos

"The idea of biological evolution helps us understand the complexity of living beings; that of chaos helps make sense of irregular time evolutions such as those of a turbulent fluid."

and further concepts yet to be developed. In this respect it is encouraging that some experts can make useful guesses about the probable evolution of the stock market, or the outcome of a political election.

To take a concrete example, instead of asking when biological evolution decided that terrestrial vertebrates have four limbs, we might ask what the probability was that at some prescribed earlier time, say the end of the Cambrian, they would develop six limbs. More precisely, can one hope for a reasonable description of life at the end of the Cambrian, such that the probability of the number of limbs of future terrestrial vertebrates could be estimated in a stable manner? Animals with six limbs would be interesting because they could adapt two limbs for manipulative purposes (centaurs) or for flying (dragons). But as centaurs and dragons do not exist, it is tempting to say that they must have had negligible probability of coming into being, and dismiss the question. One cannot so easily dismiss the problem of how life on Earth would have evolved if the great cataclysm and extinction at the end of the Cretaceous had not taken place. Are such problems outside the scope of scientific investigation? At the very least, their solution would require far greater understanding of biological evolution.

Questions of hypothetical history, or historical probabilities, are often unanswerable, at least at present. Yet these questions are not meaningless. We might one day make sense of some probabilities that arise in the history of life on Earth or in other areas. After all, such probabilities can already be estimated in the simple cases of weather prediction and astronomy of the Solar System.

David Ruelle is honorary professor at the Institut des Hautes Etudes Scientifiques, 91440 Bures sur Yvette, France.

FURTHER READING

Ruelle, D. *Chance and Chaos* (Princeton Univ. Press, Princeton, 1991; also in Penguin, 1993). Cvitanovic, P. *Universality in Chaos* (Adam Hilger, Bristol, 1984). Hao, B.-L. *Chaos II* (World Scientific, Singapore, 1990). Laskar, J. Large scale chaos and marginal stability in the solar system. In *Proc. XIth ICMP Meeting*, Paris, July 1994, 75–120 (International Press, Boston, 1995); *Celestial Mechanics* **64**,115–162 (1995).



A vision from hypothetical history?

Wiping out dirty displays

Jos van Haaren

The manufacture of liquid crystal displays still involves a surprisingly low-tech and messy process: rubbing polymer films with a velvet cloth. A twenty-year search for a cleaner alternative may finally be over.

Liquid crystal displays (LCDs) are commonly used in the screens of notebook computers and thin desktop monitors. From humble beginnings in the 1970s, modern LCD screens now have more than one million pixels, and the advent of new mobile and palm-sized devices means the market for applications is still growing. LCDs are manufactured in high-quality clean rooms (Fig. 1), with a large number of robot-driven processes. These high-tech factories require the sort of investment usually reserved for the semiconductor industry.

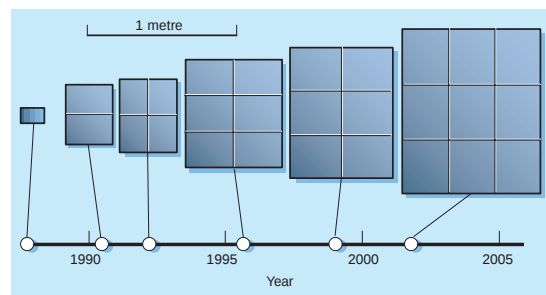
But deep inside the LCD factories, an old, mysterious and dirty process continues to be used to control the orientation of the liquid crystal molecules, and hence the clarity of the displays. At a crucial point in the production, a velvet cloth wrapped around a roller is rubbed across a surface that in turn aligns the crystals. Alternative alignment processes have been suggested, but none has been successfully transferred to large-scale manufacturing. Now, on page 56 of this issue¹, a team at IBM led by P. Chaudhari proposes to replace the velvet-covered roller with a low-energy beam of ions.

Some basics about LCDs help to paint the full picture. A liquid crystal display consists of two glass substrates with a gap 5 μm wide between them. One of the glass substrates carries an array of thin-film transistors that make it possible to control individual pixels in the display. Each pixel has a small, transparent electrode connected to the output of the transistor. The transistor inputs are connected to a data bus line, which is shared by other pixels in the same column. During operation, external integrated circuits send pulses of information through the data buses to individual transistors. A display using this technology is known as an active matrix LCD. To produce full-colour displays the pixels are grouped into threes and the outer glass substrate is covered with dots of red, green and blue filters, topped off with a transparent electrode. The space between the substrates is filled with the liquid crystal material. The transparent electrodes at the substrates serve to apply an electric field across the liquid-crystal layer, and the response of the liquid crystal translates electronic data into an image.

Liquid crystals form an intermediate phase between disordered liquids and crystalline solids. Their appearance in a bottle is



Figure 1 Clean-room operator holding a glass substrate with six liquid crystal displays (above). Over the years, there has been an increase in both the size of the displays and the number of displays manufacturers can make from a single substrate (right). This has led to an increase in productivity and a reduction in cost price. Sizes of glass substrates that are now being designed are about 1 m². This calls for better processes for, among other things, liquid-crystal alignment. At present, manufacturers use an anachronistic method to align liquid molecules: rubbing polymer films with a velvet cloth. The ion-beam technology developed by Chaudhari and co-workers¹ offers an attractive alternative.



that of a viscous, turbid fluid (Fig. 2a, overleaf). Like in a liquid, individual molecules rotate relatively easily, and a liquid crystal will completely fill the gap between two glass substrates. Liquid crystals also have some solid-like properties, such as natural ordering, so that the molecules within a small volume all point in the same direction (Fig. 2b). The direction of the molecules can be influenced by the walls of the container or by an electric field. Changing the direction of the molecules also changes the optical and electrical properties of the liquid-crystal layer. So applying a small voltage across the liquid-crystal layer in an LCD will change the amount of light transmitted by the display.

The dual nature of liquid crystals is essential for the operation of displays, but it also creates technological challenges. The liquid crystal molecules must all have the same, controlled orientation to produce a uniform

display with high contrast. If an LCD television screen were the size of a football field, the pixels would be the size of a banknote, and the liquid-crystal layer would be only 1.5 mm thick. It is not trivial to align the liquid crystals in each pixel over such a large area. To produce a high-contrast display, individual molecules have to be tilted relative to the plane of the substrate.

The rubbing process used by manufacturers to produce a defect-free display dates back to at least 1925 (ref. 2) and the earliest studies of liquid-crystal properties. All LCDs on the market today have thin polymer films that cover the side of the glass substrate in contact with the liquid crystal, and so influence its orientation. The direction of the polymer chains in these films is defined by rubbing the surface with a velvet-like cloth³. This process is tolerated because it aligns the polymers well enough, but it introduces

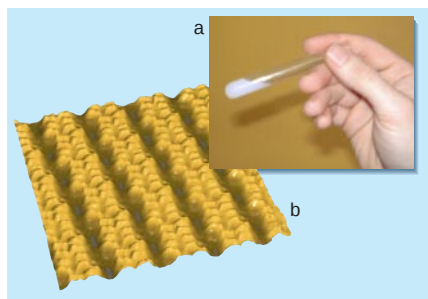


Figure 2 The dual nature of liquid crystals. a, In a bottle the liquid-crystal material looks turbid, owing to the imperfect alignment of the liquid crystal. b, Like a solid, liquid crystals can also show surprising amounts of order. Perfect alignment of individual liquid crystal molecules (an alkylated cyanobiphenyl) on a graphite surface, as observed with a scanning tunnelling microscope (J. Gerritsen, Univ. Nijmegen, The Netherlands; see also ref. 8).

debris, making it incompatible with a clean room environment. Rubbing can also leave streaks and produce electrostatic charge, which degrade image quality. Manufacturers go to great lengths and expense to avoid any sort of contamination, so a contact-free process would be a better long-term solution.

Chaudhari and co-workers¹ have developed a cleaner and more reliable process for aligning liquid crystals. They have replaced the polymer film with a thin, transparent inorganic material, known as diamond-like carbon. Exposing materials like diamond-like carbon or amorphous silicon to a low-energy beam of ions causes a rearrangement of the atoms on the surface. Placing these atoms in contact with a layer of liquid-crystal material causes the liquid crystals to align in one direction. By changing the energy and the angle of incidence of the ion beam, Chaudhari and colleagues are able to vary the tilt angle of the liquid crystal between 0° and 10°.

The IBM team have previously used ion beams to align polymer films without rubbing⁴, but now they can do away with the polymer entirely. They show that light transmission through LCDs containing diamond-like carbon films was typically 97% of that of polymer-based displays. They have also manufactured a laptop with an LCD using the diamond-like carbon film as the alignment layer. Chaudhari and his team have also shown that ion-beam alignment can be used to make monitor displays that have good image quality under different viewing angles. They use a metal mask to selectively overwrite parts of a prealigned film with the ion beam, thereby creating a two-domain display with better contrast at oblique viewing angles.

Other manufacturers are likely to follow up the IBM work with their own studies. Replacing rubbing by ion-beam alignment is an attractive idea, but the costs of this new

technology in terms of processing time and equipment need to be studied. Processing reliability, product durability and image quality will need to be investigated. Chaudhari and co-workers have paid attention to this in their prototype displays, but an industrial release of the technology is a separate and significant task. An alternative to rubbing that has not yet made it into factories is using light to align the polymer films^{5,6}. Another option developed by Fujitsu⁷ uses a type of liquid crystal that is aligned perpendicular to the substrate, rather than tilted. This configuration is radically different from conventional LCDs and leads to new issues in manufacturing and display operation.

Monitors with LCDs are currently more expensive than those with bulky cathode-ray-tubes. New manufacturing methods could bring down the costs, opening up the huge but highly price-competitive television market to LCDs. The main efficiency

improvements in the past decade have come from LCD production on larger and larger glass substrates (Fig. 1) and from changes in factory layouts. A rubbing machine for even larger substrates will be a nightmare for engineers to build and operate. Contact-free alignment, whether using ion beams or another technique, is a way to avoid the rubbing process on this scale.

Jos van Haaren is at the Philips Research Laboratories, Eindhoven, The Netherlands.
e-mail: j.a.m.m.van.haaren@philips.com

1. Chaudhari, P. *et al. Nature* **411**, 56–59 (2001).
2. Zocher, H. *Naturwissenschaften* **13**, 1015–1021 (1925).
3. van Aarle, N. J. A. M. & Tol, A. J. W. *Macromolecules* **27**, 6520–6526 (1994).
4. Chaudhari, P. *et al. Jpn J. Appl. Phys.* **37**, L55–L56 (1998).
5. Gibbons, W. *Nature* **352**, 49–50 (1991).
6. Schadt, M. H., Seiberle, H. & Schuster, A. *Nature* **381**, 212–215 (1996).
7. Ohmura, K., Kataoka, S., Sasaki, T. & Koike, Y. *SID Int. Symp. Digest of Technical Papers XXVIII* 845–848 (Society of Information Display, San Jose, CA, 1997).
8. Rastegar, A., Škarabot, M., Blij, B. & Rasing, Th. *J. Appl. Phys.* **89**, 960–964 (2001).

Protein interactions

Unspinning the web

Jeff Hasty and James J. Collins

A large-scale study of the protein network in yeast cells demonstrates the merit of taking an integrated approach to cellular dynamics, and shows the value of databases.

In the climactic scene of the movie *Independence Day*, a massive alien spacecraft, hovering just above the Earth, appears to be immune to the petty assaults mounted by the earthlings. In a last-ditch effort, the character played by Randy Quaid decides to fly his jet fighter on a kamikaze mission into the spacecraft's primary weapon. It turns out that the primary weapon is a highly connected node in the spacecraft's defence architecture. So although Quaid's attack constitutes only a pinprick, it induces an avalanche effect through the defence network, leading to the ultimate annihilation of the spacecraft.

On page 41 of this issue¹, Jeong *et al.* show that protein networks in yeast cells have wiring characteristics that are analogous to those of the alien spacecraft. These characteristics include both high resistance to random assaults, or mutations, and vulnerability to targeted attacks on specific, highly connected nodes in the network. Their data support the idea that tolerance to mutations, which has been linked to genetic redundancy, is also derived from the organization of interactions and topological position of individual proteins.

Jeong *et al.* derive their results by cleverly combining information from several different databases. First, by using data on protein–protein interactions in yeast^{2,3}, they show that the associated network follows a

power-law distribution; that is, the system contains a large number of proteins with a small number of connections and a small number of proteins with many connections. This type of network architecture, which is common to other complex systems including the Internet⁴ and metabolic networks⁵, should be both error-tolerant and vulnerable to attack⁶. Jeong and colleagues demonstrate that these properties do indeed exist by using protein-deletion data⁷ to show that the connectivity of a protein in the network is directly correlated with the likelihood that its removal will be lethal to the cell. For instance, they show that roughly two-thirds of proteins that have more than 15 connections are essential, in the sense that deleting them is lethal, whereas only one-fifth of proteins with five or fewer connections are essential.

These findings are, in many ways, intuitive. One would expect, for example, that the removal of a highly connected node in a complex network would be especially disruptive to network function. Likewise, it has been pointed out⁸ that the protein product of the *p53* tumour-suppressor gene is one of the most highly connected proteins in human cells and that mutations of *p53* can have severe consequences on basic cellular functions. Jeong *et al.* quantify this effect on a larger scale, and show how the topology of a cellular network can be related to biologi-

cal function. Their work demonstrates the value of taking an integrated approach to cellular dynamics, and shows that even though the cell is a complex, hierarchical system, it is possible to gain insight into its functional organization using relatively simple analyses.

The stage is set for further investigation of these types of protein network. How, for example, did they evolve? Possible clues can be found in related work on the emergence of power-law distributions^{6,4,9}. Such distributions will emerge if the probability that a particular node makes future connections is proportional to the number of current connections. Put another way, highly connected nodes tend to become even more connected as time goes by. What, therefore, is happening at the level of protein interactions? Certain highly connected proteins could have a special structure that enables them to bind to many different types of protein, including new ones that arise through mutation. So it may be that the proteins that make up the highly connected nodes in cellular networks share common structural features.

Modelling work¹⁰ has shown how power-law networks can arise from simple dynamical rules on the basis of evolutionary principles. This latter study demonstrates that power-law connectivity is a property of networks that are in a state of transitory expansion, suggesting that the connectivity properties of a network are a signature of its particular evolutionary state. If this is correct, it implies the existence of networks that do not follow power laws. For example, the modelling work indicates that newly developed networks are, by nature, sparsely connected and are best described by exponential distributions. Moreover, models of this type may also be relevant to the observed power-law distributions of protein family sizes¹¹, in which a 'family' consists of proteins that share sequence similarity and have similar biological functions. In this context, a power-law distribution implies the existence of 'mega-families' composed of a large number of proteins that are both structurally and functionally similar.

From a biomedical standpoint, Jeong and colleagues' findings¹ suggest that it may be unwise to select a highly connected protein as a drug target, given that inactivation of the protein could prove to be fatal or highly disruptive to the cell. Accordingly, a better strategy may be to target a less well connected protein that has a similar function. In this regard, understanding the connectivity of the network could provide likely candidates through the principle of 'guilt by association' — that is, if two proteins interact with one another, they are probably involved in similar cellular functions^{2,12}.

Jeff Hasty and James J. Collins are at the Center for BioDynamics and the Department of Biomedical

Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA.

e-mails: hasty@bu.edu

collins@bu.edu

1. Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. *Nature* **411**, 41–42 (2001).
2. Uetz, P. et al. *Nature* **403**, 623–627 (2000).
3. Xenarios, I. et al. *Nucleic Acids Res.* **28**, 289–291 (2000).
4. Barabási, A.-L. & Albert, R. *Science* **286**, 509–512 (1999).

5. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. *Nature* **407**, 651–654 (2000).
6. Albert, R., Jeong, H. & Barabási, A.-L. *Nature* **401**, 130–131 (1999).
7. Costanzo, M. C. et al. *Nucleic Acids Res.* **28**, 73–76 (2000).
8. Vogelstein, B., Lane, D. & Levine, A. J. *Nature* **408**, 307–310 (2000).
9. Simon, H. A. *Models of Man* 145 (Wiley, New York, 1957).
10. Slanina, F. & Kotrla, M. *Phys. Rev. E* **62**, 6170–6177 (2000).
11. Bader, J. S. <http://xxx.lanl.gov/abs/physics/9908032>
12. Oliver, S. *Nature* **403**, 601–603 (2000).

Astronomy

A new twist on neutron stars

Chris Fryer and Stan Woosley

Theory suggests that neutron stars should be born rotating rapidly, but in reality they spin more slowly. New calculations suggest that they may be slowed by the emission of exotic gravity waves.

A neutron star is like a gigantic atomic nucleus, packing more than a solar mass of neutrons inside a ball just 20 kilometres across. Neutron stars are born when the iron core of a massive star collapses violently inside a supernova¹. Before they collapse, the

inner cores of massive stars can have high angular momentum. Indeed theory suggests that neutron stars could be born rotating at near the maximum value they can endure without flying apart, 1,000 times per second. This is much faster than the spin rates

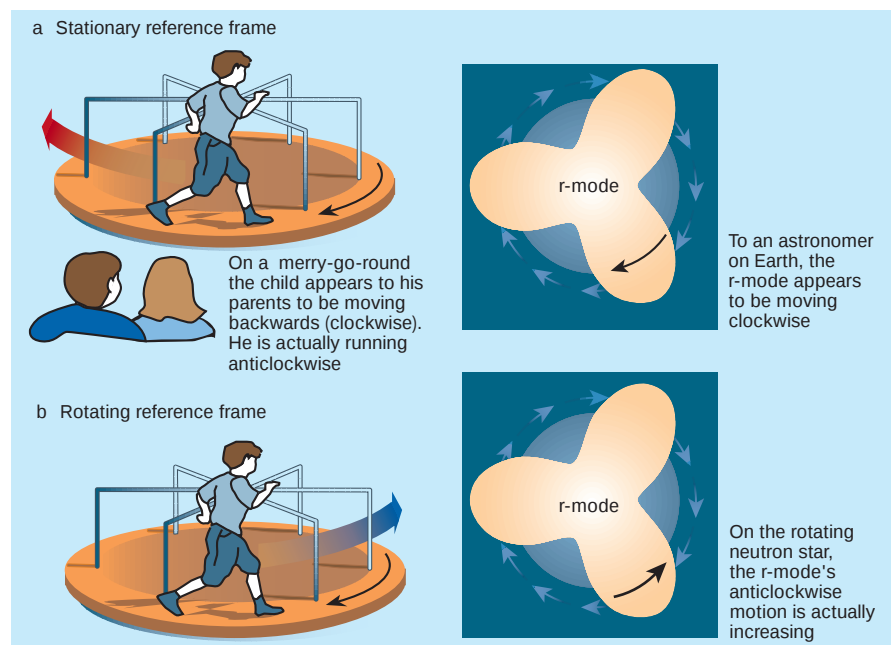


Figure 1 Rotating neutron stars and gravity waves. The growth of r-modes can be understood through an analogy with a child running anticlockwise on a clockwise-spinning merry-go-round. a, If the child is running slower than the merry-go-round is spinning, he appears to his parents sitting on a nearby bench to be moving clockwise (although slower than the merry-go-round). In the case of a spinning neutron star, an r-mode moving slowly in the opposite direction to the spinning neutron star will appear to an observer on Earth in the 'stationary reference frame' to be moving with the spin of the neutron star. The gravitational waves created by the current variations of the r-mode carry away angular momentum, causing the r-mode to slow down in the observer's stationary frame. b, In the rotating reference frame of the merry-go-round, the child is running in the opposite direction to that of the spinning merry-go-round. When the child runs faster in this frame, he appears to slow down in the stationary reference frame of his parents. Similarly, when the r-mode emits gravitational waves and slows down in the stationary reference frame of the observer, the r-mode velocity in the rotating reference frame of the neutron star is actually accelerating, causing the r-mode amplitude to grow. As the amplitude grows, more gravitational waves are emitted, leading to a runaway process. New calculations^{2,3} show that this process may be responsible for slowing down rapidly spinning neutron stars.

observed for rotating neutron stars — also known as pulsars — and indicates that angular momentum must have been removed somewhere along the way. But when, how and by how much? In two papers appearing in *Physical Review Letters*, Stergioulas and Font² and Lindblom *et al.*³ provide support for the idea that neutron stars are slowed, at least in part, by the emission of gravity waves.

Single pulsars, typically observed 1,000 years or more after their birth, rotate relatively slowly, just 30 times per second for the pulsar in the Crab Nebula. Although faster, literally, than the blink of an eye, this is a snail's pace compared with the maximum value. So why do pulsars rotate at this slower, seemingly arbitrary value? Two possibilities emerge. Either they were born rotating slowly, or they were born rotating much faster but radiated away their excess rotational energy shortly after birth. The key issues here are the spin frequency of a neutron star when it is first born, and whether that neutron star will emit copious gravitational radiation.

Theory is, at the moment, unable to discriminate between these two possibilities. For a neutron star to be born rotating slowly, angular momentum must be removed from the spinning iron core before it collapses. Calculations⁴ of the effects that might transport angular momentum in massive stars — friction, convection, circulation and the like, but ignoring magnetic fields — still allow a neutron star to be born spinning at nearly 1,000 times per second. Other calculations that crudely attempt to include the effects of magnetic forces⁵ predict, with great uncertainty, a larger effect and a very slowly rotating neutron star.

Given this uncertainty, astronomers have assumed values of the spin frequency appropriate to their tastes. Those needing enough rotation to explain gamma-ray bursts by the creation of disks around collapsed stellar cores⁶ or very luminous pulsars⁷ assume high spin frequencies of around 1,000 times per second. So do those who want rotation to be important in supernova models, especially those desirous of the large gravitational radiation signature that comes from a highly deformed, rapidly rotating neutron star. On the other hand, those seeking a simpler life and not wanting to include rotation in their supernova models adopt the slow value.

The alternative explanation for slow spin frequencies is that the neutron star is in fact born rotating with a spin frequency near 1,000 times per second, but radiates away the excess rotational energy as light or gravity waves. If this energy were emitted through a pulsar mechanism, it would drastically affect the explosion energy and emission from the supernova remnant. Observations of supernovae and young supernova remnants seem to rule this option out. Although not all astronomers agree⁷, it seems that, if neutron stars born rotating rapidly are to be slowed,

gravitational radiation must do the job. Gravitational waves are ripples in the fabric of space-time that were predicted by Einstein's theory of general relativity but that are normally too weak to be detected directly. Could young neutron stars be spinning fast enough to produce detectable gravity waves? And will such radiation significantly slow the rotation of the star?

This is where the latest results^{2,3} come in. They provide the first three-dimensional simulations of a process responsible for generating gravity waves in neutron stars, and show that it can become strong enough to slow the rotation. A rapidly rotating neutron star produces fluid motions or currents, called r-modes, which are similar to hurricanes or ocean currents on the Earth. These currents produce gravitational waves. The strength of the gravitational radiation depends on the maximum amplitude of the r-mode, which in turn depends on the competition between viscous effects that damp the flow, and driving forces that strive to increase it. The r-modes in neutron stars are driven to larger amplitudes by the same gravitational waves that they produce.

How does this positive feedback work? One way to view it is to consider an r-mode

propagating on the surface of the neutron star in a direction opposite to the neutron star's spin (Fig. 1). If the r-mode is moving more slowly than this spin, it will appear to an astronomer observing from a safe distance above the star surface to move in the same direction as the neutron star (albeit slower). The gravitational waves produced by the velocity perturbations of the r-mode carry angular momentum away from the r-modes and, to the stationary astronomer, the emission of these gravitational waves appears to make the r-mode move slower. But on the rotating neutron star the already negative velocity of the r-mode becomes more negative and its magnitude (that is, its velocity amplitude) grows.

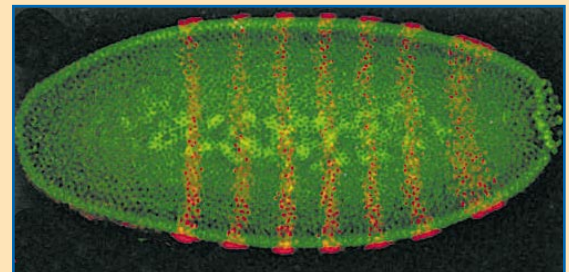
As the amplitude of the r-mode grows, its gravitational radiation increases, causing a runaway instability. Andersson⁸ and Friedman and Morsink⁹ first showed by calculation that this instability does indeed occur in neutron stars. How much it actually grows depends on the angular momentum lost through gravitational radiation, which increases rapidly with neutron star spin, and the strength of the opposing viscous forces. The viscosity is highly dependent on the temperature of the neutron star, reaching

Cell biology

Asymmetry in action

The tops and bottoms of cells generally differ in shape and structure, as well as in their protein components. In embryos, such asymmetry is crucial in establishing the pattern of cells and tissues that will make up the adult organism. One way in which the asymmetric distribution of proteins is achieved is through their encoding messenger RNAs. Take, for example, the early fruitfly embryo, pictured here. Some mRNAs (red) form stripes along the top (apical part) of the monolayer of cells that constitutes this stage of development. Others are found at the bottom. Two groups, writing in *Cell*, now delve deeper into the asymmetric distribution of RNA in fruitflies.

Andrew J. Simmonds and colleagues (*Cell* **105**, 197–207; 2001) show that the mRNA encoding a signalling protein, *Wingless*, is concentrated at the apical side of certain cells. *Wingless* is involved in many



embryonic patterning processes, and, as Simmonds *et al.* show, if its mRNA is not distributed correctly the protein does not function properly. The authors also identify the 'address labels' in the mRNA that are crucial for its distribution.

Meanwhile, Gavin S. Wilkie and Ilan Davis (**105**, 209–219; 2001) have traced the journey of some apical mRNAs from where they are produced to their final destination. They find, first, that all mRNAs diffuse at random from their source in the nucleus. In other words, apical mRNAs do not necessarily leave the nucleus on the apical side.

Instead, once outside the nucleus, particles containing the apical mRNAs make their way rapidly towards the apical part of the cell. These particles are now transported specifically in the right direction — rather than, for example, diffusing randomly throughout the cell and becoming anchored only in the apical part. They are probably transported along microtubule-based tracks, with a motor protein known as dynein acting as the transport vehicle. It remains to be seen, however, how the address labels in the mRNAs connect with this transport machinery. **Amanda Tromans**

a minimum just above 1 billion kelvin. So r-modes are most important for rapidly spinning, hot neutron stars. Such conditions could be produced either by the formation of a young hot neutron star in a supernova explosion, or in binary systems in which the neutron star accretes material from a binary companion. If r-mode amplitudes become large under these conditions, the neutron star will emit nearly all of its rotational energy as gravitational waves.

The large r-mode amplitudes predicted by Stergioulas and Font² and Lindblom *et al.*³ would make supernovae a strong source of gravitational radiation, and would provide a way of removing 90% of the rotational energy found in newly formed neutron stars within hours of the explosion that formed them. These numerical calculations give us our first glimpse of how much r-mode amplitudes can grow when they become unstable. Some uncertainties remain over the strength of the damping viscosity¹⁰, but the calculations show that the r-mode instability and its accompanying gravitational radiation must

be included in models of neutron stars. The large amplitude of the instability also implies that there is a potentially detectable gravitational radiation signal, a boon for those seeking to observe gravity waves.

Chris Fryer is in the Department of Theoretical Astrophysics, Los Alamos National Laboratory, New Mexico 87545, USA.

Stan Woosley is in the Department of Astronomy and Astrophysics, University of California at Santa Cruz, Santa Cruz, California 95064, USA.

e-mails: clf@trinary.lanl.gov

woosley@ucolick.org

1. Burrows, A. *Nature* **403**, 727–733 (2000).
2. Stergioulas, N. & Font, J. A. *Phys. Rev. Lett.* **86**, 1148–1151 (2001).
3. Lindblom, L., Tohline, J. E. & Vallisneri, M. *Phys. Rev. Lett.* **86**, 1152–1155 (2001).
4. Heger, A., Langer, N. & Woosley, S. E. *Astrophys. J.* **528**, 368–396 (2000).
5. Spruit, H. C. & Phinney, E. S. *Nature* **393**, 139–141 (1998).
6. MacFadyen, A. & Woosley, S. E. *Astrophys. J.* **524**, 262–289 (1999).
7. Wheeler, J. C., Yi, I., Hoefflich, P. & Wang, L. *Astrophys. J.* **537**, 810–823 (2000).
8. Andersson, N. *Astrophys. J.* **502**, 708–713 (1998).
9. Friedman, J. L. & Morsink, S. M. *Astrophys. J.* **502**, 714–720 (1998).
10. Jones, P. B. *Phys. Rev. Lett.* **86**, 1384–1387 (2001).

Evolutionary biology

Hybrid costs avoided

Dennis Hasselquist

The offspring of a mating between two different species are often infertile, or almost so. But it seems that some birds can avoid this apparent cost of hybridization.

The question of how new species arise has fascinated evolutionary biologists since Darwin's time. How do species diverge, and what happens when two recently diverged species come into contact, with the risk that they will mate, producing 'hybrid' offspring? In vertebrates, speciation is rarely observed — even rapid speciation is likely to proceed too slowly to be studied in a researcher's lifetime, let alone in the period

covered by a normal grant. By contrast, the effects of hybridization can be investigated in wild vertebrates, but such studies are rare. Hence the importance of the work of Veen and colleagues, who, as they describe on page 45 of this issue¹, have been looking at hybridization between two closely related bird species — pied and collared flycatchers. Their study is significant because it is based on a unique and impressively large

data set, taken from two different sites (in Sweden and the Czech Republic) over 20 years. This wealth of information allows the authors to unravel some interesting ways in which the birds avoid the apparent costs of hybridization.

When individuals from two different species mate, the offspring may be viable but the effects on their ability to reproduce (and on viability in later generations) are usually severe². There are exceptions, such as Darwin's finches on the Galapagos Islands³, but the birds studied by Veen *et al.*¹ certainly suffer adverse consequences. When pied and collared flycatchers mate with each other, the first-generation female offspring tend to be almost completely sterile. (The fitness of first-generation males, by contrast, is similar or only slightly less than that of 'pure' males from each species.) Given these facts, one might predict that heterospecific pairs (that is, pairs that consist of a pied and a collared flycatcher) would have greatly reduced reproductive output in the long term. However, Veen *et al.*'s sophisticated analyses¹ reveal some fascinating, and unexpected, patterns.

Males of the two species are distinctly different in songs, calls and coloration (Fig. 1), so it is presumably easy for females to tell them apart⁴. Despite this, pairings between pied males and collared females were much more common than would be expected by chance — particularly given that collared flycatchers outnumber their pied counterparts by 9 to 1 in the sites studied. One might think that the costs of heterospecific pairings, in terms of reproductive output, would be great, but Veen *et al.* show that it was not disadvantageous for collared females to choose pied males. The apparent costs of such pairings are lessened in several ways.

First, heterospecific pairs produced more fledglings than pure collared pairs late in the breeding season. Like many birds, pure collared flycatcher pairs fare best if they mate as early as possible in the season. But pairs of collared females and pied males have peak performance later on. All of this suggests that late-arriving collared females might do best to mate with pied males.

Second, for some heterospecific pairs, not all offspring are actually hybrids. Veen *et al.* find that many of the young raised by collared females with pied males were in fact sired by collared males. Whether this is because 'extra-pair' mating is more common for heterospecific pairs, or because the sperm of collared males win out over the sperm of pied males, remains unknown. Finally, collared females that had mated with pied males produced significantly more male than female offspring. In other words, the balance was skewed in favour of the sex of hybrid that suffered fewer effects. These three mechanisms cancelled out the apparent detrimental effects of hybridization between collared females and pied males. Indeed, the



Figure 1 Mingling without costs: collared (left) and pied flycatchers.

authors' models predict that, at some times in the breeding season, mating with a pied male may actually be the best choice for a collared female — and the data confirm this¹.

Veen *et al.*'s results show that, in vertebrates, sophisticated mechanisms may have developed to counteract the negative consequences of apparent hybridization. Such mechanisms might evolve rapidly in a location where two related species overlap. Alternatively, it is possible that these mechanisms did not evolve to cope with hybridization, but rather are a side-effect of existing female preferences.

For example, in the collared flycatcher, males that have a large white patch on the forehead are more successful in siring extra-pair young⁵. In other words, genetically speaking, it is better for collared females to mate with collared males with large forehead patches. However, socially it might be better for them to form pair bonds with collared males with slightly smaller patches if those males had better territories, and then to compensate for this by extra-pair mating. This might also explain why collared females occasionally (particularly late in the season) form pair bonds with pied males, as these males might provide better territories. As pied males have smaller forehead patches than collared males (Fig. 1), collared females in heterospecific pairs might then apply their normal mate-choice rule and engage in extra-pair matings with collared males.

Although not explicitly discussed by Veen *et al.*¹, their study has implications for speciation. Hybridization can oppose the effects of speciation — two closely related species that hybridize may eventually merge into a single

species. But it may also enhance species divergence. For example, the negative outcomes of hybridization may put pressure on the two species to evolve better ways of distinguishing between each other⁶. This would increase the differences between the species. Alternatively, the mixing of genes that occurs during hybridization might create new combinations of genes (from gene combinations that have already been shaped by considerable evolutionary pressure in the parental species), allowing a third species to evolve^{3,7}. Regardless of which of these mechanisms applies, the longer two species come into contact, mixing and exchanging genes, the more likely it is that rapid and major evolutionary changes will occur².

It is not yet known how important hybridization is to the speciation of birds. But, with the rapid development of new molecular methods for analysing DNA⁸, studies of the histories of closely related species are now possible⁹. Such studies might yield astonishing new insights into speciation, in particular revealing the involvement of hybridization. ■

Dennis Hasselquist is in the Department of Animal Ecology, Ecology Building, Lund University, 223 62 Lund, Sweden.

e-mail: dennis.hasselquist@zoekol.lu.se

1. Veen, T. *et al.* *Nature* **411**, 45–50 (2001).
2. Dobzhansky T. *Am. Nat.* **74**, 312–321 (1940).
3. Grant, P. R. & Grant, B. R. *Science* **256**, 193–197 (1990).
4. Saetre, G.-P. *et al.* *Nature* **387**, 589–592 (1997).
5. Sheldon, B. C. & Ellegren, H. *Anim. Behav.* **57**, 285–298 (1999).
6. Liou, L. W. & Price, T. *Evolution* **48**, 1451–1459 (1994).
7. Barton, N. H. & Hewitt, G. M. *Annu. Rev. Ecol. Syst.* **16**, 113–148 (1985).
8. Baker, A. J. *Molecular Methods in Ecology* (Blackwell, Oxford, 2000).
9. Andersson, M. *Proc. R. Soc. Lond. B* **266**, 1579–1585 (2000).

Neurobiology

Dopamine receptors get a boost

Francis J. White

A protein that controls the growth and survival of neurons is now shown to have another task: boosting the expression of a molecule that allows neurons to respond to the neurotransmitter dopamine.

Nerve cells in many parts of the brain communicate using the neurotransmitter dopamine, and dopamine-dependent neuronal pathways are thought to be defective in several brain disorders, including Parkinson's disease, schizophrenia and drug addiction. Much of the diversity in dopamine's effects can be explained by the fact that it works through five different types of receptor molecule. Of these, the D₁ and D₂ receptors are the most common and have been investigated most thoroughly, but the D₃ receptor, too, has received much attention since its discovery just over a decade ago¹. A few definitive roles for this receptor have now emerged^{2–4}. It is unusual

in being expressed in just a few brain regions. On page 86 of this issue, Guillin and colleagues⁵ provide evidence for another odd characteristic. They find that expression of the D₃ receptor is regulated by brain-derived neurotrophic factor (BDNF) — a protein that was once thought to be needed simply for the proliferation, maturation and survival of neurons.

Guillin *et al.*⁵ looked first at the expression of D₃ receptors in rats that have been experimentally altered to provide a 'model' of Parkinson's disease. In these rats, dopamine-releasing neurons on one side of the midbrain are destroyed by infusion of a chemical, 6-hydroxydopamine. The out-



100 YEARS AGO

The improvement in distance over which it is possible to signal has been very marked. The empirical law put forward by Mr. Marconi that, other things being equal, the distance over which signalling would be possible was proportional to the product of the heights of the masts at the two ends seems to be fairly well established as a working rule. But the improvements in transmitting and receiving apparatus have been so great that it is now possible to signal over much greater distances with the same heights of masts than was the case in 1898. For example, in 1898 Mr. Marconi was only able to cover 15 miles with vertical wires 120 ft. high, whereas to-day, according to the recent announcement made by Prof. Fleming, a distance of 200 miles from the Lizard to St. Catherine's, Isle of Wight, has been signalled over with masts only 160 ft. high... across land such great distances have not been attained, but here again we think the credit of having signalled over the greatest distance must be given to Mr. Marconi, who established in 1899 communication between Dovercourt and Chelmsford, a distance of more than 40 miles.

From *Nature* 2 May 1901.

50 YEARS AGO

It is in the tradition of British natural history that the only monograph on the water-mites of this country — the Ray Society's three volumes on "The British Hydracarina" — should have been written by two amateur naturalists, C. D. Soar and William Williamson. Williamson, born in Leith in June 1869, was during practically the whole of his working life a clerk, and latterly chief clerk, in the Scottish American Mortgage Company in Edinburgh. His interest in natural history began fortuitously, on account of his discovery, in the course of miscellaneous reading, of a series of articles describing British water-mites which appeared in *Science Gossip* in 1899 and 1900. The author was C. D. Soar, and Williamson, finding that he could easily identify from the detailed descriptions the water-mites he began to collect, got in touch with Soar, so commencing a friendship which lasted until Soar's death, almost forty years later. The contact was in a way a turning-point in his life, for the enthusiasm of one amateur stimulated the other, and Williamson's spare time now became devoted to collecting, identifying and recording hydrachnids.

From *Nature* 5 May 1951.

come is that the rats show many of the symptoms of humans with Parkinson's disease.

The D_3 receptor is normally expressed largely in an area of the brain called the nucleus accumbens, particularly in the so-called shell, which forms the deepest portions of this area. Guillin *et al.* find that the destruction of dopamine-releasing neurons in the rat model of Parkinson's disease reduces the amount of D_3 -receptor messenger RNA in neurons in the nucleus accumbens shell, and increases the amount of mRNA encoding a BDNF receptor called TrkB. Receptor levels generally increase when levels of their binding partner decrease. The authors also show that infusion of BDNF into the shell restores the expression of D_3 receptors. The implication is that the reduced expression of D_3 receptors is caused by a reduction in BDNF levels, rather than a loss of dopamine.

To find out how BDNF might be involved in regulating D_3 receptors throughout development, the authors then used mice that had been engineered to lack functional BDNF. They show that this protein is needed for the normal increase in D_3 -receptor expression that occurs shortly after birth. They also find that BDNF regulates D_3 receptors but not D_1 or D_2 receptors. Moreover, BDNF shows remarkable regional specificity: in both the BDNF-deficient mice and the rat model of Parkinson's disease, expression of D_3 receptors within the islands of Calleja — which lie beneath the shell and express the highest concentration of D_3 receptors — was normal.

Neurons produce dopamine from a precursor compound called levodopa (which is in fact the most frequently prescribed treatment for Parkinson's disease). As mentioned above, in the rat model of Parkinson's disease, dopamine-releasing neurons on only one side of the brain are damaged. When these rats are injected with levodopa, the expression of D_3 receptors in the nucleus accumbens shell increases. The animals also continuously turn away from the side of the damage, because the injection of levodopa causes dopamine receptors on the side of the damage to be activated more than those on the other side. With repeated injections, this behaviour becomes more pronounced (sensitizes), and increased expression of D_3 receptors occurs not only in the nucleus accumbens shell, but also within the region of the dorsal striatum in which the dopamine-releasing neurons are also damaged³. This suggests that the induction of D_3 -receptor expression might be responsible for behavioural sensitization to levodopa.

Guillin *et al.*⁵ investigated the involvement of BDNF in this response to levodopa. They infused soluble, antibody-labelled TrkB, which blocks BDNF, into the rats' striatum during repeated treatment with levodopa. They found that both the expres-

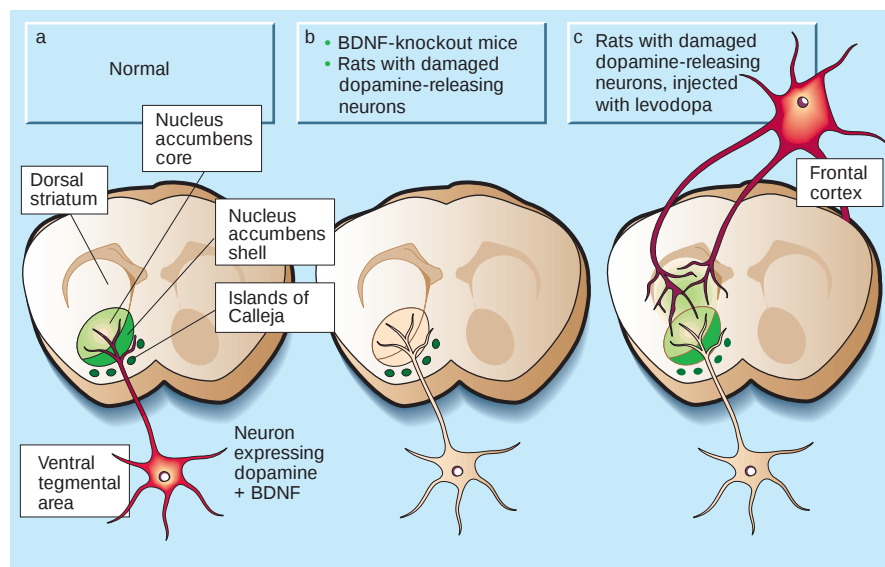


Figure 1 A neuronal growth factor broadens its scope. Brain-derived neurotrophic factor (BDNF) was once thought to be needed only for the proliferation and survival of neurons, but more of its functions have since been discovered. Guillin *et al.*⁵ now reveal that BDNF regulates the expression of the dopamine D_3 receptor. **a**, In the brains of normal rats and mice, D_3 receptors (green shading) are expressed in the shell of the nucleus accumbens region and the islands of Calleja (and, to a lesser extent, in the nucleus accumbens core). Neurons from the ventral tegmental area (VTA), which produce both dopamine and BDNF, connect to neurons in the nucleus accumbens. **b**, As Guillin *et al.* discover, in mice that lack BDNF or in rats in which dopamine-releasing neurons in the VTA are damaged, D_3 receptors are not expressed in the nucleus accumbens shell, but remain in the islands of Calleja. **c**, When the rats with damaged dopamine-releasing neurons are injected repeatedly with levodopa (the precursor of dopamine), D_3 receptors are expressed in both the dorsal striatum and the nucleus accumbens; expression of the BDNF receptor TrkB is simultaneously increased (not shown). This requires the release of BDNF from frontal-cortex neurons that express the D_1 receptor.

sion of D_3 receptors and sensitized movements were reduced. This suggests that BDNF is required for sensitization, presumably by inducing the expression of D_3 receptors. The source of the BDNF was the frontal cortex — removal of this area impaired D_3 -receptor expression and sensitization in response to levodopa.

So the authors turned their attention to the frontal cortex. They find that a single injection of levodopa induces the expression of BDNF mRNA in the frontal cortex, mainly on the side of the brain in which dopamine-releasing cells had been damaged. Moreover, the authors then show that levodopa works through D_1 or D_5 receptors in the frontal cortex to induce the expression of BDNF there. Finally, Guillin *et al.* show that, in response to repeated administration of levodopa, TrkB is expressed in the striatum even more than in the rats that are not injected with levodopa. The implication is that levodopa — which is converted to dopamine — activates D_1 or D_5 receptors in neurons of the frontal cortex. These activated receptors then stimulate the production of BDNF, which in turn acts on its receptor (TrkB) in striatal neurons to lead to the increased expression of D_3 receptors there (Fig. 1).

This comprehensive set of studies has

revealed a new function for BDNF: regulating the expression of dopamine D_3 receptors by certain neurons. In essence, BDNF thereby controls the responsiveness of those neurons to dopamine, and so is involved in long-lasting neuronal adaptations in dopamine-dependent pathways. Given that these pathways have so many roles in neurological and psychiatric disorders, the implications of these findings may be extensive. Guillin *et al.* speculate compellingly about how the regulation of D_3 receptors by BDNF could be involved in the effects of levodopa on people with Parkinson's disease, and in processes that might underlie drug addiction. Such speculation seems warranted because molecules that bind preferentially to D_3 receptors are effective in treating Parkinson's disease⁶ and in reducing cocaine-seeking behaviour in animal models of cocaine addiction⁴.

Interestingly, another neurotrophic factor — astrocytic basic fibroblast growth factor — has also been implicated in the development of behavioural sensitization, in this case to psychostimulant drugs such as cocaine^{7,8}. Continued study of the interactions between neurotrophic factors and the dopamine system will surely reveal much about how drugs affect the brain.

Francis J. White is in the Department of Cellular

and Molecular Pharmacology, Finch University of Health Sciences, The Chicago Medical School, 3333 Green Bay Road, North Chicago, Illinois 60064, USA.

e-mail: francis.white@finchcms.edu

1. Sokoloff, P., Giros, B., Martres, M.-P., Bouthenet, M.-L. & Schwartz, J.-C. *Nature* **347**, 146–151 (1990).
2. Xu, M., Koeltzow, T. E., Cooper, D. C., Tonegawa, S. & White, F.

3. J. Synapse **31**, 210–215 (1999).
3. Bordet, R. et al. *Proc. Natl Acad. Sci. USA* **94**, 3363–3367 (1997).
4. Pilla, M. et al. *Nature* **400**, 371–375 (1999).
5. Guillin, O. et al. *Nature* **411**, 86–89 (2001).
6. Bennett, J. P. Jr & Piercy, M. F. *J. Neurol. Sci.* **163**, 25–31 (1999).
7. Flores, C., Rodaros, D. & Stewart, J. J. *Neurosci.* **18**, 9547–9555 (1998).
8. Flores, C., Samaha, A. N. & Stewart, J. J. *Neurosci.* **20**, NIL7–NIL11 (2000).

Earth science

Hard-cored continents

Andrew A. Nyblade

Each continent contains pockets of ancient crust that appear to have been unaffected by tectonic forces since they formed billions of years ago. Why? There's now a fresh twist on the usual explanation.

The existence of the small, ancient cores of continents, known as cratons, has long been a puzzle. Cratons were created during the Archaean eon, more than 2.5 billion years ago, and form the oldest parts of Earth's tectonic plates. Yet they have somehow remained largely unmodified by tectonic forces. By contrast, younger parts of the continents bear the geological scars of repeated tectonic buffeting, and appear to be weaker and less stable. So, why are cratons tectonically stable? Lee and co-workers (page 69 of this issue¹) provide new insight into this question.

Not much is known about the processes that formed cratons in the Archaean. But it

has been suspected for some time that their tectonic longevity derives from 'keels' — as on sailing vessels — that extend deep into the Earth (Fig. 1). These keels are made of lithospheric mantle more than 2.5 billion years old and more than 200 km deep. The lithosphere is Earth's outermost rigid layer, and consists of the crust and uppermost mantle. The chemical composition of craton keels is thought to stem from their depletion of the basaltic constituents (Al_2O_3 , FeO , CaO) and volatile molecules (H_2O , CO_2) compared with the 'fertile' mantle that is the source of basaltic volcanism along mid-ocean ridges^{2,3}.

According to theory, a combination of the loss of basalt and volatiles makes the

keels strong enough to resist wholesale destruction by tectonic forces. This is because extraction of basaltic constituents during volcanism removes iron from the remaining mantle, making it more buoyant than its surroundings. In addition, removal of volatiles from the mantle during mantle melting increases the melting temperature and stiffness of the remaining material, making it even more resistant to tectonic forces.

Lee et al.¹ show that depletion of basaltic constituents does indeed influence the strength of lithospheric mantle, mainly by controlling the thickness to which the keel can grow. But they find that the degree of depletion is not always a function of age. Their evidence is geochemical, and comes from two regions of southwestern United States where small pieces of lithospheric mantle, called xenoliths, have been brought rapidly to the surface by volcanoes. One location is in the southern Basin and Range province, where crust 2.0–2.6 billion years in age is being deformed by tectonic forces. The other is in the Colorado plateau, a tectonically stable region of crust 1.6–2.0 billion years old that borders the Basin and Range province to the east (see map on page 70.)

By measuring the abundance of rhenium and osmium isotopes in the xenoliths, Lee and co-workers show that the lithospheric mantle beneath the sampling localities is similar in age to the overlying crust. The bulk composition of the xenoliths, together with the pressure and temperature conditions under which some of their component minerals formed, also reveal that the lithospheric mantle beneath the Basin and Range province is thinner and less depleted of basaltic constituents — that is, more fertile — than it is beneath the Colorado plateau. So it seems that it is not the age of the lithospheric mantle that correlates with its tectonic stability. Rather, given the greater basaltic depletion and thicker lithospheric mantle found under the younger yet more stable crust of the Colorado plateau, it is depletion and in turn thickness that are the determining factors.

The authors next turn to the question of how this loss of basalt controls the thickness of the lithosphere. Here they draw upon the observation that cratonic keels must in fact be neutrally buoyant, even though they contain less basalt, because they are not associated with significant perturbations in Earth's gravity field. According to the isopycnic (equal-density) condition proposed by Jordan², the negative buoyancy resulting

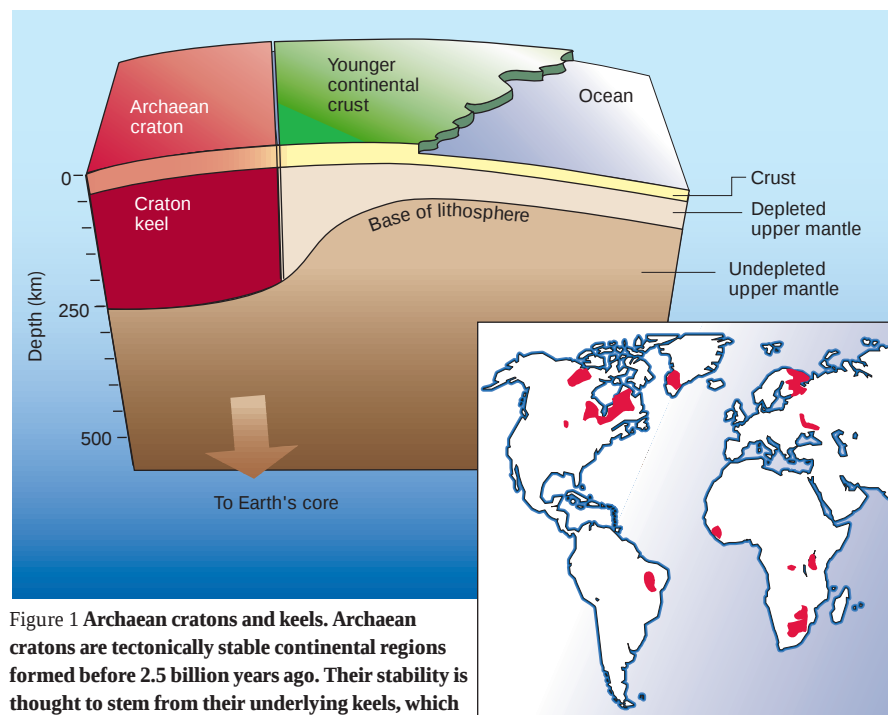


Figure 1 Archaean cratons and keels. Archaean cratons are tectonically stable continental regions formed before 2.5 billion years ago. Their stability is thought to stem from their underlying keels, which are composed of lithospheric mantle depleted of basaltic components and are at least twice as thick as the lithospheric mantle beneath younger parts of the continents and oceans. As reported by Lee et al.¹, the thickness of the keel is controlled by the degree of basalt depletion in the lithospheric mantle. The inset map shows the global distribution of Archaean cratons. (Main graphic modified from ref. 2.)

from basalt depletion is offset by the positive buoyancy that results from the lower temperatures in the keel relative to those in the surrounding, convecting mantle.

Lee *et al.*¹ compare estimated variations in the thickness, temperature and density of the lithospheric mantle with values calculated by assuming the isopycnic condition, and find that the condition holds for the southwestern United States. So, if the lithospheric mantle were to thicken further by conductive cooling in their study area, the deepest parts of the mantle lithosphere would become less buoyant and sink into the convecting mantle — as long as chemical depletion did not offset the increase in negative thermal buoyancy. In this way, the degree of basalt depletion modulates the thickness of the lithospheric mantle.

A further implication of the results¹ is that more continental crust may have formed before 2.5 billion years ago than is indicated by the present distribution of cratons. The

presence of thin Archaean lithospheric mantle beneath the Basin and Range province raises the possibility that there could be similar lithospheric mantle under other regions of the continents. We simply may not have recognized it because Archaean lithospheric mantle is commonly, but mistakenly, assumed to be thick and strong. Alternatively, the weakness of this thin lithosphere may have led to its preferential destruction through tectonic recycling. In either case, much more continental crust could have formed in the Archaean than today's distribution of cratons would indicate.

Andrew A. Nyblade is in the Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA.
e-mail: andy@geosc.psu.edu

1. Lee, C. Y., Yin, Q., Rudnick, R. L. & Jacobsen, S. B. *Nature* **411**, 69–73 (2001).
2. Jordan, T. H. *J. Petrol.* Special lithosphere issue 11–37 (1988).
3. Pollack, H. N. *Earth Planet. Sci. Lett.* **80**, 175–182 (1986).

Fundamental physics

Resistance of a perfect wire

Albert M. Chang

Intuition tells us that a wire without defects should have zero resistance. But in the real world all conductors, however perfect, have some resistance. A new study confirms that electrical contacts are the problem.

It is the dream of physicists and electrical engineers alike to build electronic devices that can conduct electrical currents with the minimal amount of resistance. The miniaturization of electronics will soon lead to devices of the smallest possible physical dimensions, in which quantum effects become important. In this context, the ultimate conductor would be a very thin one-dimensional wire that has no defects to inhibit resistance-free currents. Electrons in such a wire are ballistic — that is, the wire is so clean that the distance travelled by the electrons between collisions is longer than the wire itself. On page 51 of this issue de Picciotto *et al.*¹ describe electrical conduction in a nearly perfect, ballistic one-dimensional wire. This groundbreaking work helps to establish fundamental limits on the current-carrying capacity of ideal wires and their connections.

Imagine a perfect, extremely thin, straight wire in which electrons are allowed to move only along the wire. A perfect wire has no defects, kinks or obstacles other than a connection at each end to allow current to pass through an external circuit, and perhaps two probes along the wire to measure the voltage (Fig. 1a, overleaf). Will the motion of the electrons and hence conduction of electricity in this wire proceed without resistance? De Picciotto *et al.*¹ have cre-

ated this imaginary wire in the laboratory. They find that the resistance of a wire can be separated into two parts: an 'intrinsic resistance' due to the scattering of electrons by imperfections in the wire, and a 'contact resistance' associated with the connections to the external circuit. The intrinsic resistance is measured by the two voltage probes, which draw negligible current. The authors find that in their defect-free wire the intrinsic resistance does actually reach zero, although there is a finite contact resistance of around 13 k Ω .

The vanishing of the intrinsic resistance agrees with the simple notion that the current-carrying electrons should move freely if there are no obstacles. Our picture of electrical resistance as the result of momentum-changing deflections on charge-carriers dates back to the work of Drude around 1900. The most effective deflections are those that scatter charge-carriers in the opposite direction to that of their flow. In the late 1950s, Landauer² became fascinated with the idea of electronic miniaturization and proposed a conceptual framework for understanding electrical conduction in one-dimensional wires. Landauer realized that the wire can be thought of as being connected to electrochemical potential reservoirs, in which many electrons of different energies are available for conduction

(Fig. 1b). Once an electron enters the wire it cannot change energy, and only momentum changes can affect the electrical current. So where the wire is connected to the reservoirs there is a mismatch of energies and a contact resistance develops. Conversely, in a regular three-dimensional wire, the contact resistance is very small and so tends to go unnoticed. But in a one-dimensional conductor, with no impurities, the intrinsic resistance must vanish although the contact resistance is high.

Until the 1980s most current measurements on mesoscopic devices (typically a micrometre or less in size) used a two-terminal geometry, in which voltage difference is measured solely between the current source and drain. But after Webb *et al.*³ developed the technology to place multiple terminals on small metallic wires and rings, it was discovered that the four-terminal — two voltage probes in addition to two current leads — resistance of a non-ballistic wire changes if the direction of an external magnetic field is reversed. This unexpected result was explained by Buttiker⁴, who generalized Landauer's ideas to devices with multiple terminals. He pointed out that, in the absence of magnetic impurities, a four-terminal resistance should be unchanged by swapping over the current and voltage probes and reversing the magnetic field.

After considering the multiple-terminal case, Buttiker proposed that a two-terminal resistance must always contain a contact resistance, even if the wire is ballistic. This contact resistance can be measured directly between point contacts that are shorter than the mean free path of electrons in a non-ballistic conductor^{5,6}. Two-terminal measurements of ballistic wires also found their resistance to be quite large, around 13 k Ω . Is this resistance just the sum of the contact resistances?

De Picciotto *et al.*¹ have built a multi-terminal one-dimensional conductor to address this issue. Classically a wire is considered to be one-dimensional if its width exactly accommodates the size of the charge carriers with no room for wiggling. Electrons are essentially point objects, so this is technically unfeasible. But this is where quantum mechanics comes to the rescue. If an electron is made to occupy the lowest quantum-mechanical energy state in the lateral directions, without access to higher excited states, it would be free to move only in one dimension. The key is to confine electrons so tightly that the energy levels of the excited states are too high for them to reach. One way of doing this is to cool the wire to sufficiently low temperatures — but not superconducting temperatures — so that the excited states are thermally inaccessible and conduction is strictly one dimensional.

To create the perfect one-dimensional

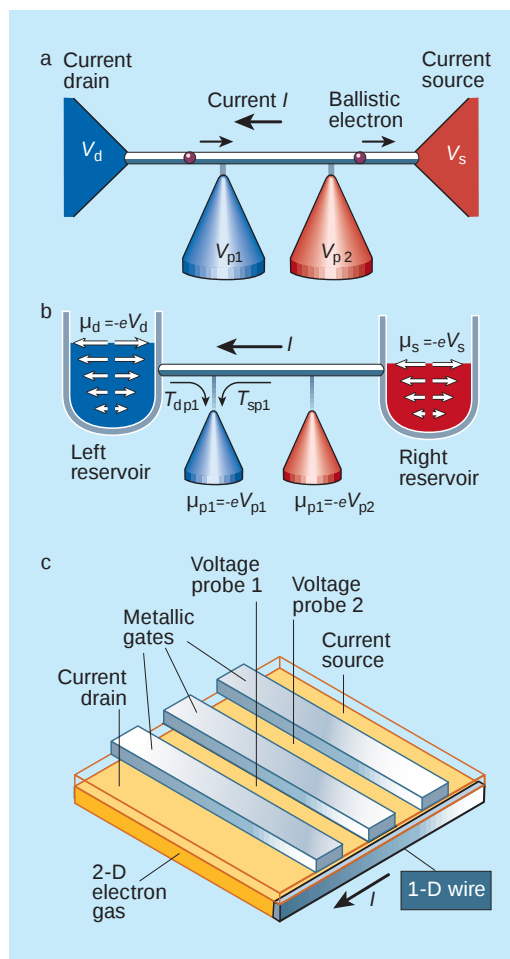


Figure 1 Measuring resistance in one-dimensional wires. a, An ideal one-dimensional conductor with two current terminals and two voltage probes, which draw no current. The two terminal resistance, R_{2p} , is defined as $(V_s - V_d)/I$, whereas the four-terminal intrinsic resistance, R_{in} , is given by $(V_{p2} - V_{p1})/I$. b, Landauer's concept² of a perfect one-dimensional wire, in which electron reservoirs with unbalanced electrochemical potentials ($\mu = \mu_s - \mu_d$) provide current source and drain. Note that $\mu = -eV$, where μ is the electrochemical potential, e is the fundamental electric charge and V is the voltage. The measured resistances depend on the transmission and reflection probabilities of the electrons within the conductor. In particular, for a four-terminal intrinsic resistance to be zero, the transmission and reflection probabilities from the left and right reservoirs into a voltage probe must be nearly equal ($T_{sp1} = T_{dp1}$). This is only possible if the conductor is ballistic and devoid of obstacles. c, The perfect one-dimensional wire created by de Picciotto *et al.*¹ with two current and two voltage terminals created from the same two-dimensional electron gas.

wire, de Picciotto *et al.* use a precise semiconductor growth technique called cleaved edge overgrowth⁷ to grow alternate crystalline layers of GaAs and AlGaAs in orthogonal directions, which form a neat edge. Even more remarkably they have succeeded in attaching non-invasive voltage probes to the minuscule wire. The GaAs/AlGaAs sheet forms a two-dimensional electron gas, and the authors place metallic electrodes on the surface to isolate the one-dimensional wire from the rest of the sheet (Fig. 1c). When negative voltages are applied to these electrodes they deplete the two-dimensional electron gas beneath them but preserve the one-dimensional wire along the edge. The width of the metallic electrodes defines the length of the isolated wire. The separation between the electrodes also creates two narrow strips in the two-dimensional electron gas that act as voltage probes.

This set-up allows both four-terminal and two-terminal measurements to be made on the same wire. In order to measure the intrinsic resistance, it is essential that the voltage probes do not disturb the current flow. With a one-dimensional wire this is usually impossible, but in their system de Picciotto *et al.* can tune the effect of the voltage probe until it is almost non-existent. They show that whereas the intrinsic resis-

tance measured between the non-invasive voltage probes vanishes, the two-terminal resistance of about 13 k Ω can be directly attributed to the current contacts. This non-zero contact resistance indicates that there is a minimum resistance for the passage of electricity regardless of how perfect a wire may be.

It would be interesting to explore the possibility of circumventing the limitations imposed by the presence of contact resistances. One can imagine measuring the electrical properties of one-dimensional conductors using a contact-free method, such as capacitance coupling, to induce charge motion. The results of studies like that of de Picciotto and colleagues will apply equally to carbon nanotubes and other one-dimensional systems, and will become increasingly important as electronics gets ever smaller.

Albert M. Chang is in the Department of Physics, Purdue University, West Lafayette, Indiana 47907-1396, USA.

e-mail: yingshe@physics.purdue.edu

1. de Picciotto, R., Stormer, H. L., Pfeiffer, L. N., Baldwin, K. W. & West, K. W. *Nature* **411**, 51–54 (2001).
2. Landauer, R. *IBM J. Res. Dev.* **1**, 223–231 (1957).
3. Webb, R. A., Washburn, S., Umbach, C. P. & Laibowitz, R. B. *Phys. Rev. Lett.* **54**, 2696–2699 (1985).
4. Buttiker, M. *Phys. Rev. Lett.* **57**, 1761–1764 (1986).
5. van Wees, B. J. *et al.* *Phys. Rev. Lett.* **60**, 848–850 (1988).
6. Wharam, D. A. *et al.* *J. Phys. C* **21**, L209–L214 (1988).
7. Pfeiffer, L. N. *et al.* *Appl. Phys. Lett.* **56**, 1697–1699 (1990).

Daedalus

Scroll-reading

The ancient world guards its secrets well. In the buried ruins of Herculaneum, the many scrolls that make up the library of Philodemus wait to be deciphered. Sadly, the scrolls have been carbonized, and it is extremely difficult to unwind them for normal reading. Daedalus now has an idea.

The nuclear magnetic resonance imager should make it possible to scan such scrolls without unrolling them. Interpreting a scrolled image would take very clever software, but it seems feasible. More problematically, could the best modern machines resolve the lettering? Fortunately, the basic pixel of a coil is curved, like that of a scroll, and also extends in depth, again like a scroll. With luck, a scroll rotated in a superconducting coil, and slowly moved in and out of it, should yield up its lettering to a controlling scholar.

Of course, nobody believes in the Wisdom of the Ancients any more. No matter what amazing statements or beliefs are found in antique libraries, letter files or rubbish dumps, the job of scholars will be to decipher, translate and interpret. Simple letters inviting somebody to a party, or denying an allegation, may be handled most easily; documents claiming to be part of a revelation will give a lot more trouble.

The basic problem, however, is what nucleus to look for in the magnetic-resonance output. Ancient inks were based on soot (carbon), which might have little contrast with the surrounding carbonaceous papyrus or vellum, unless it contains germanium or potassium, from coal or wood. It might be better to examine the vehicle (gum), which would have more hydrogen, and more mobile hydrogen, than the lettering itself. If so, preliminary steaming of the scroll could help. If any of the inks were based on black iron gallate, the signal from paramagnetic iron should be easy to detect. Either way, Daedalus reckons that some sort of contrast change between ink and background could be detected by the instrument and highlighted by its software, enabling a scholar to read the entire scroll without the pain of unwinding it.

Even the problem of dating the find may become tractable, if the changes in the ink, vehicle and substrate turn out to be fairly predictable in time. And if in the future a better method becomes available, why, the scroll is still there, undamaged by clumsy attempts to unroll it.

David Jones

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

Proteins are traditionally identified on the basis of their individual actions as catalysts, signalling molecules, or building blocks in cells and microorganisms. But our post-genomic view is expanding the protein's role into an element in a network of protein–protein interactions as well, in which it has a contextual or cellular function within functional modules^{1,2}. Here we provide quantitative support for this idea by demonstrating that the phenotypic consequence of a single gene deletion in the yeast *Saccharomyces cerevisiae* is affected to a large extent by the topological position of its protein product in the complex hierarchical web of molecular interactions.

The *S. cerevisiae* protein–protein interaction network we investigate has 1,870 proteins as nodes, connected by 2,240 identified direct physical interactions, and is derived from combined, non-overlapping data^{3,4}, obtained mostly by systematic two-hybrid analyses³. Owing to its size, a complete map of the network (Fig. 1a), although informative, in itself offers little insight into its large-scale characteristics. Our first goal was therefore to identify the architecture of this network, determining whether it is best described by an inherently uniform exponential topology, with proteins on average possessing the same number of links, or by a highly heterogeneous scale-free topology, in which proteins have widely different connectivities⁵.

As we show in Fig. 1b, the probability that a given yeast protein interacts with k other yeast proteins follows a power law⁵ with an exponential cut-off⁶ at $k_c \approx 20$, a topology that is also shared by the protein–protein interaction network of the bacterium *Helicobacter pylori*⁷. This indicates that the network of protein interactions in two separate organisms forms a highly inhomogeneous scale-free network in which a few highly connected proteins play a central role in mediating interactions among numerous, less connected proteins.

An important known consequence of the inhomogeneous structure is the network's simultaneous tolerance to random errors, coupled with fragility against the removal of the most connected nodes⁸. We find that random mutations in the genome of *S. cerevisiae*, modelled by the removal of randomly selected yeast proteins, do not affect the overall topology of the network. By contrast, when the most connected proteins are computationally eliminated, the network diameter increases rapidly. This simulated tolerance against random mutation is in agreement with results from systematic mutagenesis

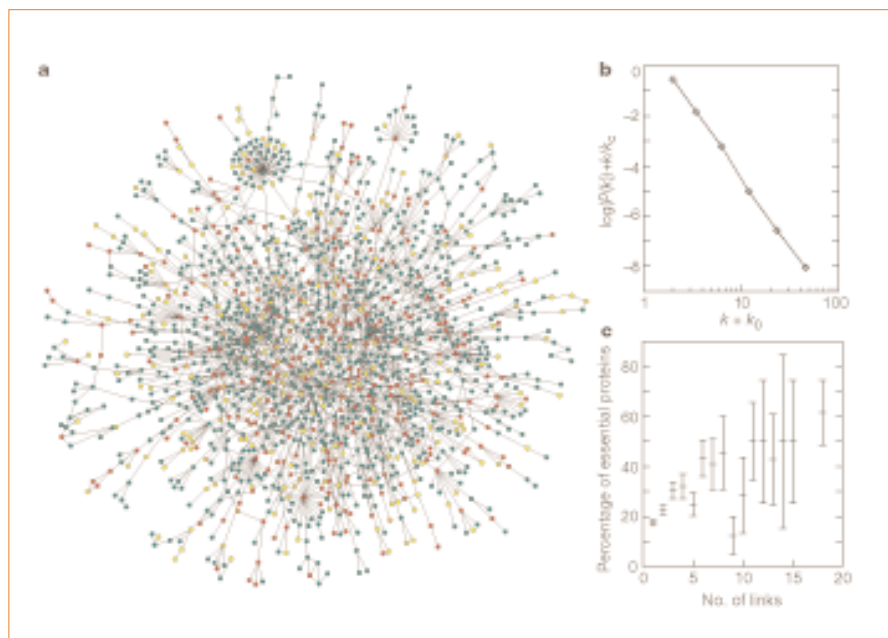


Figure 1 Characteristics of the yeast proteome. **a**, Map of protein–protein interactions. The largest cluster, which contains $\sim 78\%$ of all proteins, is shown. The colour of a node signifies the phenotypic effect of removing the corresponding protein (red, lethal; green, non-lethal; orange, slow growth; yellow, unknown). **b**, Connectivity distribution $P(k)$ of interacting yeast proteins, giving the probability that a given protein interacts with k other proteins. The exponential cut-off⁶ indicates that the number of proteins with more than 20 interactions is slightly less than expected for pure scale-free networks. In the absence of data on the link directions, all interactions have been considered as bidirectional. The parameter controlling the short-length scale correction has value $k_0 \approx 1$. **c**, The fraction of essential proteins with exactly k links versus their connectivity, k , in the yeast proteome. The list of 1,572 mutants with known phenotypic profile was obtained from the Proteome database¹³. Detailed statistical analysis, including $r = 0.75$ for Pearson's linear correlation coefficient, demonstrates a positive correlation between lethality and connectivity. For additional details, see <http://www.nd.edu/~networks/cell>.

experiments, which identified a striking capacity of yeast to tolerate the deletion of a substantial number of individual proteins from its proteome^{9,10}. However, if this is indeed due to a topological component to error tolerance, then, on average, less connected proteins should prove to be less essential than highly connected ones.

To test this, we rank-ordered all interacting proteins based on the number of links they have, and correlated this with the phenotypic effect of their individual removal from the yeast proteome. As shown in Fig. 1c, the likelihood that removal of a protein will prove lethal correlates with the number of interactions the protein has. For example, although proteins with five or fewer links constitute about 93% of the total number of proteins, we find that only about 21% of them are essential. By contrast, only some 0.7% of the yeast proteins with known phenotypic profiles have more than 15 links, but single deletion of 62% or so of these proves lethal. This implies that highly connected proteins with a central role in the network's architecture are three times more likely to be essential than proteins with only a small number of links to other proteins.

The simultaneous emergence of an inhomogeneous structure in both metabolic^{5,11} and protein interaction networks suggests that there has been evolutionary selection of a common large-scale structure of biological networks and indicates that future systematic protein–protein interaction studies in other organisms will uncover an essentially identical protein–network topology. The correlation between the connectivity and indispensability of a given protein confirms that, despite the importance of individual biochemical function and genetic redundancy, the robustness against mutations in yeast is also derived from the organization of interactions and the topological positions of individual proteins¹². A better understanding of cell dynamics and robustness will be obtained from an integrated approach that simultaneously incorporates the individual and contextual properties of all constituents in complex cellular networks.

H. Jeong*, S. P. Mason†, A.-L. Barabási*, Z. N. Oltvai†

*Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA
e-mail: alb@nd.edu, zno008@nwu.edu

†Department of Pathology, Northwestern University Medical School, Chicago, Illinois 60611, USA

1. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. *Nature* **402**, 47–52 (1999).
2. Eisenberg, D., Marcotte, E. M., Xenarios, I. & Yeates, T. O. *Nature* **405**, 823–826 (2000).
3. Uetz, P. et al. *Nature* **403**, 623–627 (2000).
4. Xenarios, I. et al. *Nucleic Acids Res.* **28**, 289–291 (2000).
5. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. *Nature* **407**, 651–654 (2000).
6. Amaral, L. A., Scala, A., Barthélemy, M. & Stanley, H. E. *Proc. Natl Acad. Sci. USA* **97**, 11149–11152 (2000).
7. Rain, J.-C. et al. *Nature* **409**, 211–215 (2001).
8. Albert, R., Jeong, H. & Barabási, A.-L. *Nature* **406**, 378–382 (2000).
9. Winzler, E. A. et al. *Science* **285**, 901–906 (1999).
10. Ross-Macdonald, P. et al. *Nature* **402**, 413–418 (1999).
11. Fell, D. A. & Wagner, A. in *Animating the Cellular Map* (eds Hofmeyr, J.-H., Rohwer, J. M. & Snoep, J. L.) 79–85 (Stellenbosch Univ. Press, 2000).
12. Wagner, A. *Nature Genet.* **24**, 355–361 (2000).
13. Costanzo, M. C. et al. *Nucleic Acids Res.* **28**, 73–76 (2000).

Cell culture

Progenitor cells from human brain after death

Culturing neural progenitor cells from the adult rodent brain has become routine^{1,2} and is also possible from human fetal tissue^{3,4}, but expansion of these cells from postnatal and adult human tissue, although preferred for ethical reasons, has encountered problems^{5–8}. Here we describe the isolation and successful propagation of neural progenitor cells from human post-mortem tissues and surgical specimens. Although the relative therapeutic merits of adult and fetal progenitor cells still need to be assessed, our results may extend the application of these progenitor cells in the treatment of neurodegenerative diseases.

As a source of neural progenitor cells, we used brain tissue from an 11-week-old post-natal male, who died of extracerebral complications of myofibromatosis, and from a resectioned temporal cortex from a 27-year-old male (courtesy of J. Alksney). The tissue was removed, sectioned 2 hours after death and placed in cold antibiotic-containing Hank's buffered-salt solution, then processed for culture 3 hours later. Representative sections of hippocampus, ventricular zone, motor cortex and corpus callosum were taken. The temporal cortex tissue was provided *en bloc* and placed in chilled Hank's buffered-salt solution and processed for culture 3 hours after removal. The adult tissues were divided into hippocampal formation, white matter and remaining cortical grey matter. Tissues were finely diced and then dissociated by enzymic digestion with papain, DNase I and neutral protease for 45 min at 37 °C, as described for rodent tissue¹.

We initially plated isolated cells onto fibronectin-coated plates in DMEM:F12 medium containing glutamine, amphotericin-B, penicillin, streptomycin and 10% fetal bovine serum. After 24 hours, the medium was replaced with DMEM:F12 supplemented with BIT-9500 (bovine serum albumin, transferrin, insulin; Stem Cell Technologies), 20 ng ml⁻¹ basic fibroblast growth factor (FGF-2), 20 ng ml⁻¹ epidermal growth factor, and 20 ng ml⁻¹ platelet-derived growth factor AB.

We had limited success using these condi-

tions, but when the medium was supplemented with 25% conditioned medium from rat stem cells that had been genetically modified to overproduce a secreted form of FGF-2 and its stem-cell cofactor, the glycosylated form of cystatin C (ref. 9), plating efficiency, as well as initial survival and growth, greatly improved. The medium was changed every two days and the cultures were replated onto twice the surface area to accommodate proliferative expansion. All tissue samples gave neural progenitor cells, but the highest yields were from hippocampus and ventricular zone. For long-term storage, cultures were

dissociated with trypsin, rinsed and cryopreserved in growth medium (without growth factors) containing 10% dimethylsulphoxide.

Cells from the 11-week-old tissue grew at log phase for more than 70 population doublings before showing signs of *in vitro* senescence (a significant reduction in growth rate). The adult tissues were expanded for more than 30 doublings before senescence (Fig. 1). Neurons were spontaneously generated at all stages in these cultures (Fig. 1a,b), and more complete differentiation could be induced by withdrawing growth factors and stimulating the cells with forskolin and retinoic acid (Fig. 1c)^{1,9}. Neonatal and adult cultures produced similar proportions of neurons and astrocytes, although the number of spontaneously generated neurons was lower than reported for fetal cultures^{3,4} and decreased significantly as cultures reached senescence. Oligodendrocytes were rare in most cultures.

So far, we have processed 23 tissue samples from people of different age groups. Most samples have yielded viable progenitor cells, with the longest post-mortem interval

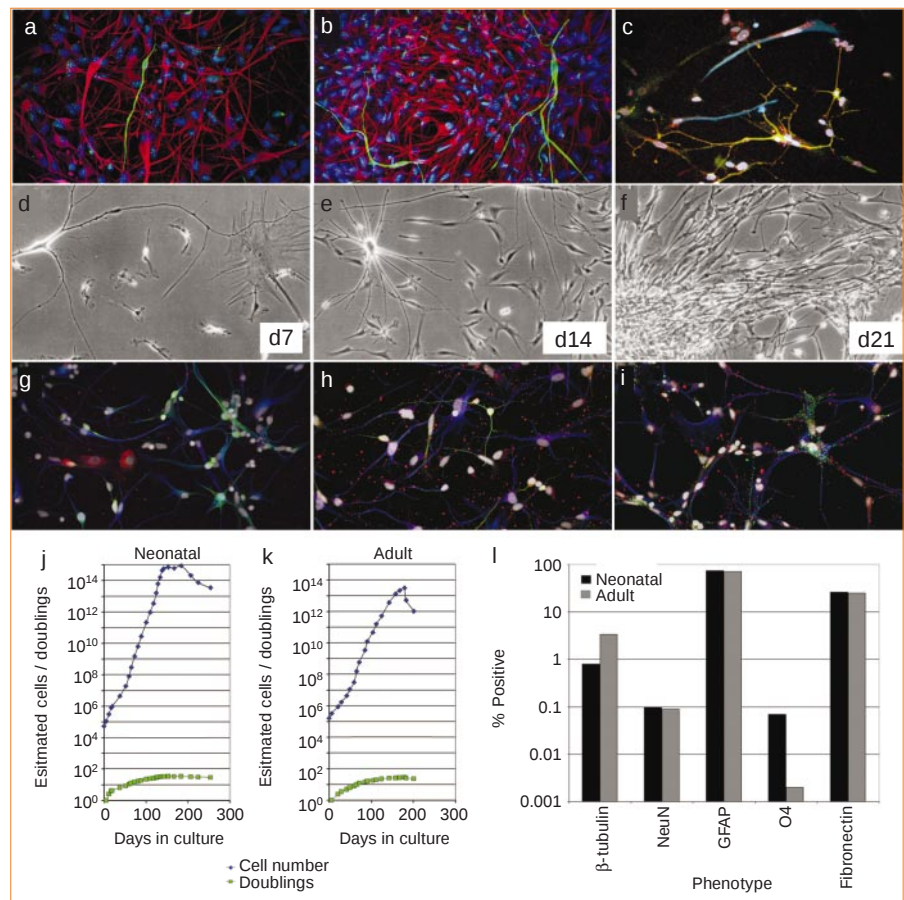


Figure 1 Neural progenitor cells isolated from postmortem human brain. **a–i**, Cells obtained from brains of two males: an 11-week-old neonate (**a–c**) and a 27-year-old adult (**d–i**). Proliferative cells (**a, b**) were stained¹ for immature neurons (β-tubulin; green), astrocytes (GFAP, glial fibrillary acidic protein; red) and nuclei (DAPI stain; blue). Differentiated cells (**c**) were stained for neurons: Map2abc (red) and neurofilament 150 (green), for GFAP (blue), and nuclei (white). Adult cells as seen by phase-contrast microscopy at days (**d**) 7, 14 and 21 after plating (**d–f**). Differentiated cells are shown in the third row: **g**, fibronectin (fibroblasts, red), β-tubulin (green), GFAP (blue), nuclei (white); **h**, Map2abc (green), GFAP (blue), nuclei (white); **i**, immature glia (A2B5, green), GFAP (blue), nuclei (white). The growth of neonatal and adult cells in culture is shown in the bottom row: **j, k**, phenotypic counts of differentiated neonatal and adult cells, respectively; **l**, immunostaining of differentiated cells: NeuN, post-mitotic neuron marker; O4, immature oligodendrocyte marker; fibronectin, connective tissue.

before culture being about 20 hours. Tissues from young individuals yield significantly more cells per gram and these cells have a higher proliferative capacity. Intrinsic differences in growth potential and/or lineage potential may affect the utility of cell transplantation or other applications for therapy. Questions relating to the *in vitro* lifespan of these cells and the role of telomerase¹⁰, as well as the importance of DNA methylation, still need to be addressed.

Using cells cultured from embryonic tissues bypasses some of these concerns, but also raises complex ethical and societal issues. Careful evaluation and consideration of the relative merits of post-mortem or adult-derived cells and fetal progenitor cells will be necessary.

Theo D. Palmer*†, **Philip H. Schwartz‡**,
Philippe Taupin*, **Brian Kaspar***,
Stuart A. Stein‡, **Fred H. Gage***

*The Salk Institute, Laboratory of Genetics, 10010
North Torrey Pines Road, La Jolla, California 92122,
USA

e-mail: gage@salk.edu

†Stanford University, Department of Neurosurgery,
MSLS P304, 1201 Welch Road,
Stanford, California 94305-5487, USA

‡Children's Hospital of Orange County, Brain and
Tissue Bank for Developmental Disorders, 455 South
Main Street, Orange, California 92668, USA

1. Palmer, T. D., Markakis, E. A., Willhoite, A. R., Safar, F. & Gage, F. H. *J. Neurosci.* **19**, 8487–8497 (1999).
2. Reynolds, B. A. & Weiss, S. *Science* **255**, 1707–1710 (1992).
3. Svendsen, C. N., Caldwell, M. A. & Ostenfeld, T. *Brain Pathol.* **9**, 499–513 (1999).
4. Fricker, R. A. *et al. J. Neurosci.* **19**, 5990–6005 (1999).
5. Pincus, D. W. *et al. Ann. Neurol.* **43**, 576–585 (1998).
6. Johansson, C. B., Svensson, M., Wallstedt, L., Janson, A. M. & Frisen, J. *Exp. Cell Res.* **253**, 733–736 (1999).
7. Pagano, S. F. *et al. Stem Cells* **18**, 295–300 (2000).
8. Roy, N. S. *et al. Nature Med.* **6**, 271–277 (2000).
9. Taupin, P. *et al. Neuron* **28**, 385–397 (2000).
10. Ostenfeld, T. *et al. Exp. Neurol.* **164**, 215–226 (2000).

Boundary effects

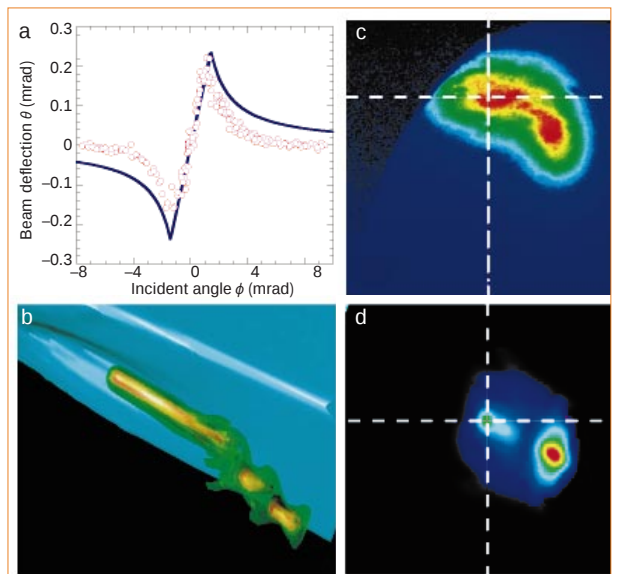
Refraction of a particle beam

The refraction of light at an interface is familiar as a rainbow or the 'bending' of a pencil in a glass of water. Here we show that particles can also be refracted and even totally internally reflected, as evidenced by an electron beam of 28.5×10^9 electron volts being deflected by more than a milliradian upon exiting a passive boundary between a plasma and a gas — the electron beam is bent away from the normal to the interface, just like light leaving a medium of higher refractive index. This phenomenon could lead to the replacement of magnetic kickers by fast optical kickers in particle accelerators, for example, or to compact magnet-less storage rings in which beams are guided by plasma fibre optics.

Refraction is caused by electrostatic plasma fields set up when plasma electrons are

Figure 1 Experimental and simulation results demonstrating refraction of an electron beam at a plasma–gas interface.

a, Actual electron-beam deflection (circles), measured using a beam-position monitor, and the theoretical deflection (blue line) as a function of the incident angle. **b**, Simulation: perspective image of a beam emerging from plasma (turquoise); the inward motion of the plasma electrons is visible as a depression in the plasma surface behind the beam. **c**, Experiment: image of the beam downstream of the plasma, showing the deflected beam and the undeflected transient (at the crosshairs); **d**, head-on view of image in **b**. The beam consisted of 1.9×10^{10} electrons at 28.5 GeV in a gaussian bunch of length $\sigma_z = 0.7$ mm and spot size $\sigma_x \approx \sigma_y \approx 40$ μm . The plasma, with radius 2.3 mm, length 1.4 m and density $1 \times 10^{14} \text{ cm}^{-3}$, was created by photoionization of lithium vapour by an ArF laser. The angle, ϕ , between the electrons' initial trajectory and the plasma boundary was controlled by adjusting the tilt angle of the final laser-beam mirror.



expelled by the collective space charge force of the head of the beam. The plasma ions in the beam path are more massive and remain, constituting a positively charged channel through which the latter part of the beam travels. The ions provide a net force that focuses the beam^{1,2}. When the beam comes close to the plasma boundary, the ion channel becomes asymmetric, producing a deflecting force in addition to the focusing force. This formation of an asymmetric plasma lens³ gives rise to the bending of the beam path at the interface.

The order of magnitude of this deflection can be estimated, yielding an expression for the deflection angle, Θ , as a function of the incident angle, ϕ . This is the effective non-linear Snell's law for the electron beam refraction, valid for ϕ greater than Θ :

$$\Theta = (8 \times N r_e) / (\pi \sqrt{2} \pi \gamma \sigma_z \sin \phi)$$

where $N/\sqrt{2} \pi \sigma_z$ is the charge per unit length of the beam, r_e is the classical electron radius, γ is the beam's energy in units of mc^2 and α is a factor of order one that is a weak function of plasma density and bunch length. When ϕ is less than Θ , this equation breaks down and the beam is internally reflected. Simulations⁴ show that $\Theta \approx \phi$ (critical reflection) for small values of ϕ .

We tested this analytical model by using the electron beam at the Stanford Linear Accelerator Center (Final Focus test facility), as described^{5,6}. Sample results are shown in Fig. 1 and compared with a full three-dimensional electromagnetic particle-in-cell computer simulation⁷. In Fig. 1a, the solid curve represents the prediction from the model (with $\alpha = 0.2$): for incident angles smaller than 1.3 mrad, the beam appears to be internally reflected, in agreement with the model.

Figure 1b shows a snapshot of the real

space of the beam and plasma electron density (turquoise) from a simulation. A transient at the head of the beam is apparent because of the finite time that it takes the plasma to respond to the beam. The tail portion is deflected towards the plasma and is near the plasma boundary. The transient results in the characteristic splitting of the beam images downstream, as shown in Fig. 1c, d.

The simulations and experimental results presented here show that it is possible to refract and even reflect a particle beam from a dilute plasma gas. Remarkably, for a 28.5-GeV beam that can bore through several millimetres of steel, the collective effects of a plasma are strong enough to 'bounce' the beam off an interface that is one million times less dense than air.

Patric Muggli*, **Seung Lee***, **Thomas Katsouleas***, **Ralph Assmann†**, **Franz-Joseph Deckert†**, **Mark J. Hogan†**, **Richard Iverson†**, **Pantaleo Raimondi†**, **Robert H. Siemann†**, **Dieter Walz†**, **Brent Blue‡**, **Christopher E. Clayton‡**, **Evan Dodd‡**, **Ricardo A. Fonseca‡**, **Roy Hemker‡**, **Chandrashekhar Joshi‡**, **Kenneth A. Marsh‡**, **Warren B. Mori‡**, **Shoquin Wang‡**

*University of Southern California, Los Angeles,
California 90089, USA

e-mail: katsoule@usc.edu

†Stanford Linear Accelerator Center, Stanford
University, Stanford, California 94309, USA

‡University of California, Los Angeles, California
90095, USA

1. Su, J. J., Katsouleas, T. & Dawson, J. M. *Phys. Rev. A* **41**, 3321–3331 (1990).
2. Whittum, D., Sessler, A. & Dawson, J. M. *Phys. Rev. Lett.* **64**, 2511–2514 (1990).
3. Chen, P. *Part. Accel.* **20**, 171–182 (1987).
4. Katsouleas, T. *et al. Nucl. Instrum. Meth. Phys. Res. A* **455**, 161–165 (2000).
5. Muggli, P. *et al. Phys. Rev. Sp. Top. AB* (submitted).
6. Hogan, M. J. *et al. Phys. Plasmas* **7**, 2241–2248 (2000).
7. Hemker, R., Mori, W. B., Lee, S. & Katsouleas, T. *Phys. Rev. Sp. Top. AB* **3**, 61301–61305 (2000).

Protein function

Chaperonin turned insect toxin

Antlions are larvae of the Myrmeleontidae family that live on other insects¹ by sucking out the body fluid from their prey, after first paralysing it with a toxin produced by salivary bacteria. Here we show that the paralysing toxin produced by bacterial endosymbionts in the saliva of *Myrmeleon bore* larvae is a homologue of GroEL, a protective heat-shock protein known as a molecular chaperone. The amino-acid residues critical for this protein's toxicity are located away from the regions essential to its protein-folding activity, indicating that the dual function of this GroEL homologue may benefit both the antlion and the endosymbiont.

The saliva of the larvae of *M. bore* (Fig. 1a), a pit-building antlion, contains the endosymbiont *Enterobacter aerogenes* (Fig. 1b), which, when grown up in culture and injected into German cockroaches (*Blattella germanica*), rapidly paralyses them. We purified one of the insecticidal proteins from culture broth and found that it migrated on a denaturing SDS–polyacrylamide gel as a single band at a position corresponding to a relative molecular mass of about 63K (Fig. 1c). Partial amino-acid sequencing of this toxin indicated that it was a GroEL homologue.

GroEL, also known as chaperonin, is a product of the *groE* gene of *Escherichia coli* and was first discovered as a mutant that inhibits bacteriophage growth². GroEL forms a homo-oligomer with a double-toroid structure and, in combination with the protein GroES and ATP, acts as a molecular chaperone to ensure correct folding and assembly of proteins^{2–4}. However, the paralytic and insecticidal activities described here cannot be explained in terms of a chaperonin function.

The GroEL homologue from *E. aerogenes* showed no acute toxicity towards mice, but it rapidly paralysed and killed cockroaches when injected at a minimum dose of 2.7 ± 1.6 ng (mean $\pm$ s.e.m.; $n = 3$); the recombinant protein encoded by its complementary DNA and expressed in *E. coli* was equally toxic. By contrast, GroEL from *E. coli* did not paralyse the insects even at doses as high as 2 μ g.

We cloned the *groE* gene from *E. aerogenes* and found that only 11 residues in the GroEL homologue had alignments different from the residues in GroEL from *E. coli* (Table 1) — except at the carboxy terminus, where the GroEL homologue has a methionine residue and is shorter than GroEL by three residues. The amino-acid residues Val 100 (valine at the 100th residue), Asn 101, Asp 338 and Ala 471 are crucial for toxicity, as shown by the marked reduction in toxicity of mutants carrying the substitutions Ile 100, Thr 101, Glu 338 and Gly 471 (Table 1). The importance of these residues was confirmed by reversing the

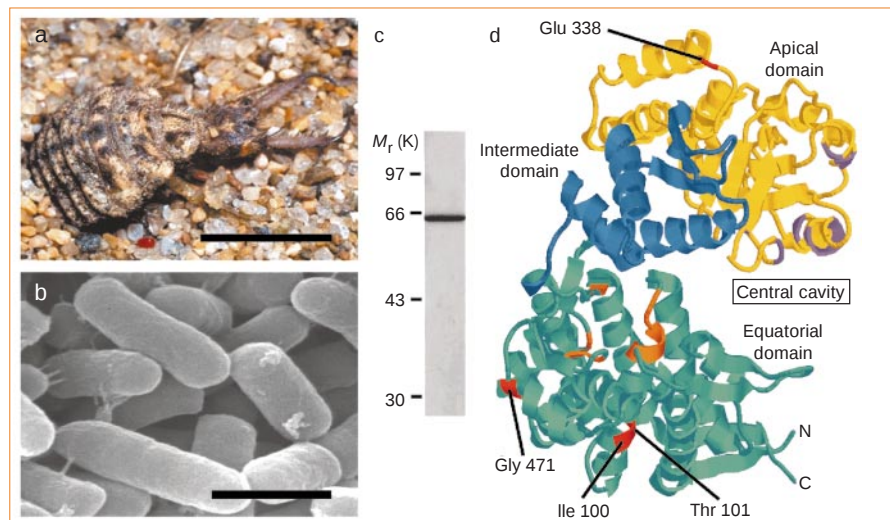


Figure 1 An insect toxin produced by a salivary endosymbiont of antlion larvae is a GroEL homologue. **a**, Larva of *M. bore*. Scale bar, 5 mm. **b**, Scanning electron micrograph of the bacterial endosymbiont *E. aerogenes*. Scale bar, 1 μ m. **c**, Electrophoresis on a 10% SDS–polyacrylamide gel of the toxic protein purified from a culture of *E. aerogenes*. **d**, The three-dimensional structure of a subunit in a 14-mer GroEL molecule from *E. coli* (created using a PDB file, 1DER). The apical, intermediate and equatorial domains are coloured yellow, blue and green, respectively. Locations of the residues in which mutation confers toxicity are shown in red. Sites at which mutation blocks the binding of polypeptide and GroES in the apical domain⁸ and the residues involved in ATP binding in the equatorial domain^{6,7} are shown in purple and orange, respectively.

individual mutations (substituting valine for isoleucine at residue 100, and so on) in *E. coli* GroEL, which conferred toxicity on the protein (Table 1).

In the crystal structure of *E. coli* GroEL^{5–7}, Glu 338 is located in the apical domain, far from the polypeptide- and GroES-binding sites^{7,8}, whereas Ile 100, Thr 101 and Gly 471 are in the equatorial domain, far from the ATP-binding pocket that faces the central cavity (Fig. 1d)^{6,7}. Thus, neither GroES binding nor ATP binding to the GroEL homologue is effective in generating toxicity. Injection of the GroES homologue from *E. aerogenes* together with ATP and the GroEL homologue scarcely influenced the minimum paralysing dose (3.9 ± 0.3 ng; $n = 3$).

Chaperonins from bacterial pathogens can function as cell-signalling molecules,

stimulating human monocytes, leukocytes, fibroblasts and epithelial cells to release pro-inflammatory cytokines⁹. The toxicity of the GroEL homologue towards insects can be seen as an effect of a bacterial extracellular chaperonin on eukaryotic cells. The homologue may act on particular receptors in insects to induce paralysis, having evolved this non-chaperone function to establish a mutually beneficial antlion–symbiont relationship. Our finding that insecticidal proteins are produced by endosymbionts to help in capturing prey is likely to extend to many other fluid-feeding carnivorous insects.

Naofumi Yoshida*, Kenji Oeda†, Eijiro Watanabe‡, Toshiyuki Mikami‡, Yoshikazu Fukita§, Keiichiro Nishimura¶, Koichiro Komai*, Kazuhiko Matsuda*

*Department of Agricultural Chemistry, Faculty of Agriculture, Kinki University, 3327-204 Nakamachi, Nara 631-8505, Japan

e-mail: kmatsuda@nara.kindai.ac.jp

†Agricultural Chemicals Research Laboratory,

§Organic Synthesis Research Laboratory, Sumitomo Chemical Co., and ‡Genomic Science Laboratories, Sumitomo Pharmaceuticals, 2-1 Takatsukasa

4-Chome, Takarazuka, Hyogo 665-8555, Japan

¶Research Institute for Advanced Science and Technology, Osaka Prefecture University,

1-2 Gakuencho, Sakai, Osaka 599-8570, Japan

1. Matsura, T. *Ecol. Res.* **1**, 15–24 (1986).
2. Zeilstra-Ryalls, J., Fayet, O. & Georgopoulos, C. *Annu. Rev. Microbiol.* **45**, 301–325 (1991).
3. Sigler, P. B. *et al.* *Annu. Rev. Biochem.* **67**, 581–608 (1998).
4. Ranson, N. A., White, H. E. & Saibil, H. R. *Biochem. J.* **333**, 233–242 (1998).
5. Braig, K. *et al.* *Nature* **371**, 578–586 (1994).
6. Boisvert, D. C., Wang, J., Otwinowski, Z., Horwich, A. L. & Sigler, P. B. *Nature Struct. Biol.* **3**, 170–177 (1996).
7. Xu, Z., Horwich, A. L. & Sigler, P. B. *Nature* **388**, 741–750 (1997).
8. Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. *Nature* **371**, 614–619 (1994).
9. Ranford, J. C., Coates, A. R. M. & Henderson, B. *Exp. Rev. Mol. Med.* <http://www-ermm.bcu.cam.ac.uk/00002015.htm> (2000).

Table 1 Paralytic activity of GroELs

<i>E. aerogenes</i> GroEL homologue	MPD (ng per insect)	<i>GroEL</i> from <i>E. coli</i>	MPD (ng per insect)
Wild type	3.3 $\pm$ 1.1	Wild type	ND
V100I	ND	I100V	26.7 $\pm$ 3.2
N101T	ND	T101N	61.3 $\pm$ 3.7
V125T	53.3 $\pm$ 4.1	T125V	ND
D338E	ND	E338D	13.6 $\pm$ 1.4
T347A	18.8 $\pm$ 2.5	A347T	ND
I426L	16.0 $\pm$ 2.7	L426I	ND
G428D	46.8 $\pm$ 4.1	D428G	ND
K430R	51.2 $\pm$ 4.4	R430K	ND
A471G	ND	G471A	14.4 $\pm$ 1.3
S527N	25.7 $\pm$ 1.7	N527S	ND
P530A	51.5 $\pm$ 2.8	A530P	ND

Mutants are represented by the usual notation, in which the wild-type amino-acid residue is given first in single-letter code, then the position in the protein sequence, and finally the substituent amino acid. Minimum paralytic dose (MPD) values are expressed as means $\pm$ s.e.m. of minimal GroEL amounts required to paralyse cockroaches within 10 min after injection ($n = 5$). ND, not determined (because of low toxicity; MPD > 1,000). Further details are available from K.M.

Hybridization and adaptive mate choice in flycatchers

Thor Veen^{*†}, Thomas Borge[‡], Simon C. Griffith^{‡§}, Glenn-Peter Sætre[‡], Stanislav Bures^{||}, Lars Gustafsson^{*} & Ben C. Sheldon[§]

^{*} Department of Animal Ecology, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden

[†] Department of Biology, University of Groningen, 9750 Haren, The Netherlands

[‡] Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden

[§] EGI, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

^{||} Palacky University, Laboratory of Ornithology, 77146 Olomouc, Czech Republic

Hybridization in natural populations is strongly selected against when hybrid offspring have reduced fitness. Here we show that, paradoxically, pairing with another species may offer the best fitness return for an individual, despite reduced fitness of hybrid offspring. Two mechanisms reduce the costs to female collared flycatchers of pairing with male pied flycatchers. A large proportion of young are sired by conspecific male collared flycatchers through extra-pair copulations, and there is a bias in favour of male offspring (which, unlike females, are fertile) within hybrid pairs. In combination with temporal variation in breeding success, these cost-reducing mechanisms yield quantitative predictions about when female collared flycatchers should accept a male pied flycatcher as a mate; empirical data agree with these predictions. Apparent hybridization may thus represent adaptive mate choice under some circumstances.

Animal and plant hybrid zones are often seen as natural laboratories in which to study evolutionary processes associated with speciation¹. One very important stage in the speciation process is the development of pre-mating isolating mechanisms between two divergent populations. Reduced fitness of hybrids is thought to select for mechanisms that reduce the likelihood of mating events between populations. Although low rates of hybridization may occur relatively frequently between closely related taxa in some groups^{1–3}, it is quite difficult to study the fitness consequences of hybrid matings in natural populations. Here we show in two closely related species of birds that, despite the presence of effective pre-mating isolating mechanisms, hybrid pairings occur at an unexpectedly high frequency. Dissection of the fitness consequences of different pairing and mating events shows, however, that hybrid pairing may be selectively favoured in a seasonally changing environment.

Rates and consequences of hybrid pairing

Pied and collared flycatchers (*Ficedula hypoleuca* and *F. albicollis* respectively) have two separate areas of sympatry (range overlap) within their European range, where hybridization occurs^{4–7}. In the isolated population breeding on the Swedish island of Gotland, collared flycatchers are numerically dominant, comprising 95% of all breeding birds. In central Europe a hybrid zone runs east–west through the Czech republic⁶, and in our study area there, collared flycatchers are again numerically dominant (85% of breeding birds). Pairing in both populations is species assortative (see Table 1 for a test of the null hypothesis of random mating: Swedish population: $\chi^2_4 = 943.1$, $P \ll 0.0001$; Czech population: $\chi^2_4 = 310.0$, $P \ll 0.0001$). Extensive life-history data collected over 20 years in the Swedish population allows the estimation of fitness of hybrids relative to the parental species (Table 2). F₁ hybrids have reduced fitness relative to the parental species, with greatly reduced fitness for F₁ hybrid females owing to almost complete sterility (Table 2). There is no recorded case where an F₁ hybrid female has recruited offspring to the breeding population, and all females for which hybrid status has been established using genetic markers (see below) have been infertile. Males have slightly reduced fertility (92% as assessed by egg hatchability) relative to males of the parental species (Table 2). Data from the Czech population suggest that hybrid

fitness is similarly, or even more strongly, reduced there⁶. Studies using genetic markers have confirmed that hybrid males are fertile (ref. 5 and B.C.S., unpublished work). Neither sex of hybrid shows any sign of reduced viability once they have reached breeding age. Male hybrids apparently suffer reduced fitness owing to lower recruitment of offspring relative to collared flycatchers, but not relative to pied flycatchers (Table 2). The low recruitment of offspring for pied flycatchers, and thus possibly for male hybrids too, may be due to reduced philopatry (return to natal site) of their young⁸. These data suggest that the relative reduction in fitness for a female that pairs with a heterospecific male will be substantial (up to 75% of fitness lost).

Given the rarity of pied flycatchers in both hybrid zones, it is not difficult to understand why a female pied flycatcher may sometimes be forced to pair with a male collared flycatcher: there may simply be no unpaired male pied flycatchers remaining^{3,9}. Males of both species apparently show no mate discrimination¹⁰. However, it is much more difficult to understand why a female collared flycatcher (the numerically dominant species) should choose to pair with a male pied flycatcher, particularly since males of the two species differ strikingly in terms of plumage, song and calls, and given that

Table 1 Mating frequencies of flycatchers (number of pairing events observed)

Swedish birds				
Female of pair				
Male of pair	Collared	Hybrid	Pied	Total
Collared	5,567	15	84	5,666
Hybrid	110	1	8	119
Pied	72	3	172	247
Total	5,749	19	264	6,032
Czech birds				
Female of pair				
Male of pair	Collared	Hybrid	Pied	Total
Collared	601	19	15	635
Hybrid	18	0	5	23
Pied	12	6	69	87
Total	631	25	89	745

experimental studies have shown perfect discrimination by females in areas of sympatry⁷. Theoretical models of mate choice suggest that the criteria determining the acceptance of a potential mate should be relaxed if mate choice occurs under time constraints¹¹, and as with many birds breeding in temperate regions, flycatchers are selected to begin breeding as early as possible^{12–15}. Female collared flycatchers pairing with male pied flycatchers breed later (2.6 days later in Sweden and 3.5 days in the Czech Republic) than females in pure collared pairs ($F_{1,5493} = 12.51$, $P = 0.0004$, and $F_{1,512} = 3.56$, $P = 0.06$ respectively, with a general linear model (GLM) controlling for year-associated variation). These observations suggest that females may choose to breed in mixed pairs because of time constraints on breeding, but it remains difficult to understand how time constraints could be so severe that females accept the substantial loss in fitness that seems to result from hybridization.

Conspecific extra-pair paternity

Long-term pedigree data from the Swedish population suggest, however, that the cost of mating with a male pied flycatcher may be reduced by extra-pair copulations with conspecific males. Of 28 offspring recruited from PF × CF (male pied flycatcher × female collared flycatcher) pairs, only 12 (43%) were identified as first-generation hybrids when first handled, the rest (57%) being identified as pure collared flycatchers (identification was made without any knowledge of the identity of an individual's parents). In contrast, 25 of 36 offspring (69%) recruiting from CF × PF pairs were identified as first-generation hybrids when first handled ($\chi^2 = 6.98$, $P < 0.01$). One interpretation of these data is that the offspring of females in PF × CF pairs are frequently sired by conspecific males. However, these data also suggest occasional errors in field identification, since 8/36 (22%) of offspring from CF × PF pairs were identified as pure collared flycatchers, which should not be possible given that intraspecific brood parasitism is rare or absent in this population¹⁶. Mis-identifications could arise in this fashion if alleles from collared flycatchers were dominant to those from pied flycatchers at loci controlling plumage patterns.

We tested whether rates of extra-pair paternity were elevated in PF × CF pairs by microsatellite genotyping families of different specific combinations from both the Swedish and Czech populations. A male's share of paternity in the brood depended on pair-type for both the Swedish population (Fig. 1a; GLM: $F_{3,41} = 12.66$, $P < 0.0001$) and the Czech population (Fig. 1a; GLM: $F_{3,45} = 5.47$, $P = 0.0027$). This effect was largely due to the higher rate of extra-pair paternity among PF × CF pairs, compared with the other three pair types (Fig. 1a). There was no difference between the two populations in either the pattern of paternity across different pair types (GLM: $F_{3,86} = 0.527$, $P = 0.67$) or the overall rate of paternity ($F_{1,86} = 0.811$, $P = 0.37$). The overall mean rate of extra-pair paternity in pure collared flycatcher families (14.5%) is similar to

a previous estimate from the Swedish population (15.5%; ref. 16), whereas the mean rate of extra-pair paternity in PF × CF pairs is approximately four times greater at 59% in the Swedish population, agreeing well with the figure of 57% suggested by the pedigree data.

A more relevant comparison, in terms of the fitness of the offspring produced, is of the proportion of young that are sired by a conspecific male (Fig. 1b). In all pure pairings ($N = 44$ CF × CF pairs and $N = 16$ PF × PF pairs), all offspring were sired by a male

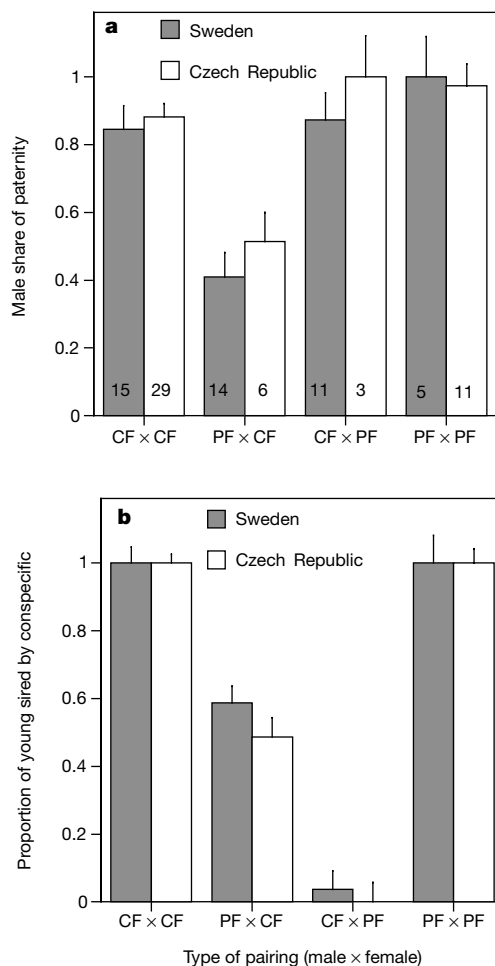


Figure 1 Male share of paternity in relation to pairing type in flycatcher families.

a. Mean (+s.e.) share of paternity for male flycatchers in relation to pairing type in two separate hybrid zones. Numbers in bars give sample size (number of broods analysed). **b.** Mean (+s.e.) proportion of young sired by a male flycatcher conspecific with the female of the pair; sample sizes as for **a**.

Table 2 Fitness components for individual collared and pied flycatchers and their hybrids in the Swedish hybrid zone

Species	Sex	Fitness component (mean ± s.e.)									
		Hatchability	N	Fledging success	N	Lifetime fledged young	N	Reproductive attempts	N	Lifetime recruitment	N
CF	Male	0.946 ± 0.003†	1,700	0.744 ± 0.006†	2,375	6.26 ± 0.09†	2,259	1.65 ± 0.02†	3,264	0.67 ± 0.02†	2,316
	Female	0.884 ± 0.005†	2,266	0.615 ± 0.007†	3,157	5.21 ± 0.08†	2,978	1.59 ± 0.02†	4,192	0.54 ± 0.02†	3,008
Hybrid	Male	0.873 ± 0.049‡	29	0.664 ± 0.051†	47	5.28 ± 0.61†	47	1.58 ± 0.10‡	79	0.33 ± 0.14‡	52
	Female	0.092 ± 0.047‡	31	0.151 ± 0.057‡	42	0.84 ± 0.31‡	38	1.70 ± 0.16‡	43	0.00 ± 0.00‡	39
PF	Male	0.952 ± 0.010†	107	0.811 ± 0.023‡	141	6.70 ± 0.36†	140	1.42 ± 0.07‡	173	0.35 ± 0.08‡	148
	Female	0.867 ± 0.024†	141	0.720 ± 0.026§	200	5.44 ± 0.27†	199	1.20 ± 0.04‡	262	0.27 ± 0.04§	212
Test statistic	Male	$F = 4.67^*$		$F = 4.94^{**}$		$F = 2.04$		$F = 4.19^*$		$F = 10.11^{***}$	
	Female	$F = 144.8^{***}$		$F = 41.04^{***}$		$F = 20.37^{***}$		$F = 29.34^{***}$		$F = 15.15^{***}$	

Hatchability and fledging success are expressed as the proportion of eggs hatching, and young fledging, relative to the number of eggs laid in the clutch. Lifetime fledged young, N reproductive attempts, and lifetime recruitment are the total number of young fledged, number of times observed breeding, and the number of offspring recruited to the breeding population, respectively, over an individual's lifetime. All statistics refer to comparisons within sexes. Values with different superscript symbols differ significantly from each other (Tukey tests, $P < 0.05$).

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

conspecific with the female of the pair. In mixed pairs, there was a highly significant difference between PF × CF and CF × PF pairs ($F_{1,32} = 35.38$, $P < 0.0001$; $N = 20$ and $N = 14$ respectively). All extra-pair sired young in PF × CF pairs were sired by a male conspecific with the female, resulting in 56% of offspring from these pairings being non-hybrid. In only one case (involving two nestlings from a brood of five in the Swedish population) were extra-pair offspring from a CF × PF pair sired by a male conspecific with the female. Pedigree data from this population also suggest a low rate of conspecific extra-pair copulation in CF × PF pairs, because 3/36 recruited offspring (8.3%) were identified as pure pied flycatchers when first recaptured. We interpret the difference in the rate of conspecific extra-pair paternity between the two types of mixed pairs to be due to the great difference in the abundance of potential conspecific extra-pair sires.

A further test of whether the rate of heterospecific fertilization differs between the two classes of mixed pairs can be made using F_1 hybrid individuals and a polymerase chain reaction (PCR)-based test which detects the presence/absence of a 32 base pair (bp) mitochondrial DNA indel, which is a fixed difference between the two species¹⁷, as this allows identification of the maternal species of the hybrid. If extra-pair copulations are more frequent in PF × CF pairs, then F_1 hybrids with collared flycatcher mothers should be under-represented relative to the proportion of PF × CF pairs comprising mixed pairings. We screened 31 adult F_1 hybrids from the two populations combined (individuals not included in the paternity analysis above), for which hybrid status was confirmed by microsatellite genotyping, and found that 9/31 (29%) had collared flycatcher mtDNA. This is almost significantly fewer than expected based on the proportions of mixed pairs of the two types in the two populations (99 CF × PF and 84 PF × CF pairs: $\chi^2_1 = 3.55$, $P = 0.059$), but close to the proportion expected (28%) if the rate of conspecific extra-pair mating is taken into account (goodness-of-fit $\chi^2_1 = 0.03$, $P = 0.87$).

The elevated rates of extra-pair paternity in PF × CF broods need not represent elevated rates of extra-pair copulation, because sperm from male collared flycatchers may have an advantage owing to conspecific sperm precedence^{18,19}. Further work is required to determine whether female collared flycatchers in mixed pairs pursue an active strategy to increase the rate of extra-pair copulation dependent on the species of the male mate. Irrespective of the behavioural mechanism, the result is that female collared flycatchers in mixed pairs produce a high proportion of non-hybrid offspring, meaning that the costs of heterospecific pairing are lower than they appear at first sight. Female collared flycatchers would be able to reduce the costs of hybridization even further by completely cuckolding their heterospecific mates, yet we found only 3/20 such families. Male collared flycatchers seem to adjust their rate of parental care in response to their perceived share of paternity^{20,21}, and it is possible that the share of paternity that male pied flycatchers gain, on average, reflects a minimum level at which a male's parental assistance will still be guaranteed. The reproductive success of unassisted female flycatchers is substantially reduced relative to those whose offspring receive paternal care²².

Sex ratio bias

The long-term data from Sweden also suggest that there may be a male bias in the sex ratio of recruits from mixed pairs, compared to pure collared pairs, because 18/28 recruits (64%) from PF × CF pairs were male, compared to 1,218/2,528 (48%) from CF × CF pairs: ($\chi^2_1 = 2.892$, $P = 0.089$). We tested whether the same bias was present at the nestling stage by using genetic markers to identify the sex of nestlings from PF × CF and CF × CF broods. Combining data for the two populations, we found that the mean brood sex ratio in PF × CF pairs was higher than in pure collared pairs (Fig. 2a; GLM: $F_{1,169} = 5.24$, $P = 0.023$), and corresponded quite well to that suggested by pedigree data. The biased sex ratio in mixed pairs

might be explained by increased mortality of females (the heterogametic sex) at any time after fertilization, in accordance with Haldane's rule²³. Surprisingly, however, long-term data from the Swedish population give no indication that either the proportion of eggs hatching, or the proportion of eggs fledging, are lower for PF × CF pairs than CF × CF pairs; in fact the reverse is true for fledging success (Fig. 2b). Both would be expected to be lowered in mixed pairs relative to pure pairs, if heterogametic female hybrids suffered higher mortality. Hence, if the male-biased sex ratio is the result of differential mortality of male and female embryos it must occur very early in development.

An alternative hypothesis is that the sex ratio bias occurs because females in PF × CF pairs adjust the sex ratio of eggs they ovulate in favour of males. Recent evidence suggests that female birds may have some degree of control over their offspring sex ratio^{24–27}. Selection for such control would be strong and consistent when one sex of offspring had a reproductive value of zero, although conditional sex allocation of this precision may seem biologically implausible. Whether the sex ratio bias is caused by biased mortality or biased production, the result is that females in mixed pairs invest less parental effort, and hence fewer resources, in female offspring with lower reproductive value. This represents a second means by which the costs of heterospecific pairing are lower than they appear at first sight.

Fitness equivalence of mixed pairing

Our data suggest that, for a female collared flycatcher, the cost of pairing with a male pied flycatcher may be reduced by two mechanisms. In contrast, the rarity of suitable conspecific mates, even as extra-pair copulation partners, means that female pied flycatchers, the rarer species, pay almost full costs of hybridization. We tested the fitness costs of heterospecific mate choice directly by assessing the number of offspring and grand-offspring recruited to the breeding population from different pairings in the Swedish population. The mean number of recruited offspring was independent

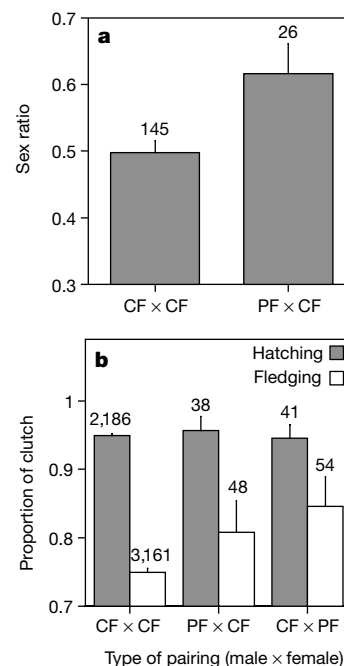


Figure 2 Effect of pairing type on sex ratio and viability of brood. **a**, Sex ratio within broods of flycatchers in relation to pairing type. Numbers above bars give sample sizes. **b**, Hatching and fledging success (mean + s.e. proportion of eggs hatched) in relation to pairing type. Numbers above bars give the number of breeding attempts in each category. Hatching success was independent of pairing type ($F_{2,3263} = 0.04$, $P = 0.92$), while there was a weak effect of pairing type on fledging success ($F_{2,3263} = 3.20$, $P = 0.04$).

of the type of pairing (Fig. 3a), suggesting that hybrid offspring do not suffer any reduction in viability compared with pure-bred offspring. However, if the cost of hybridizing is due to the production of sterile offspring, this cost should be seen in terms of the number of grand-offspring produced, which corresponds better to the reproductive value of offspring. We found a highly significant effect of pairing type on the number of grand-offspring produced (Fig. 3b). Female pied flycatchers mated to male collared flycatchers produced significantly fewer grand-offspring per breeding attempt than did either pure collared pairs or female collared flycatchers mated to male pied flycatchers; the latter two did not differ significantly.

The data on paternity and sex ratio within mixed species and pure families, together with the life-history data (Table 2) allow the calculation of the expected relative fitness of the three types of pairings (pure collared; male pied $\times$ female collared; and male collared $\times$ female pied). Assuming total sterility among female hybrids, we can calculate the expected fitness of the three types of pairings in two ways. First, discounting male fitness that is lost due only to reduced fertility suggests that the fitness of the two mixed pairings relative to a pure collared pairing will be in the ratio: $CF \times CF : PF \times CF : CF \times PF = 1.00 : 0.82 : 0.47$. This calculation allows for the possibility that the reduced recruitment success of hybrid males is due to their offspring's showing greater dispersal. Alternatively, assuming that the recruitment data in Table 2 are representative of the relative recruitment success of male collared flycatchers and hybrids, and discounting the fitness of hybrid

males accordingly, yields a predicted ratio of $CF \times CF : PF \times CF : CF \times PF = 1.00 : 0.72 : 0.27$. Both sets of predictions lie within the 95% confidence intervals of the observed fitness of the three pair types (Fig. 3b), although the prediction based on the full life-history data (that is, prediction 2) shows a closer agreement with the data.

The calculated average loss in fitness for a female collared flycatcher pairing with a male pied flycatcher, as opposed to a male collared flycatcher (18–28%), is still apparently substantial despite the cost-reducing mechanisms. However, this comparison ignores the fact that mean laying date in $PF \times CF$ pairings is 2.6 days later than in pure CF pairings. In the Swedish collared flycatcher population there is strong selection for breeding as early as possible^{12,13}, with the standardized selection gradient on laying date being $\beta = -0.343 \pm 0.035$ s.e. ($P < 0.0001$; s.d. of laying date is 4.5 days). A difference in mean laying date of 2.6 days should result in a loss of relative fitness of about 20%. However, the two pair-types differ with the respect to the seasonal decline in reproductive success (Fig. 4a), which explains why, despite breeding later, $PF \times CF$ pairs fledge a greater proportion of eggs than do pure collared pairs (Fig. 2b). Consequently, the overall reproductive success per breeding attempt does not differ between $PF \times CF$ pairs and pure collared pairs, despite the fact that the latter breed earlier. It is not clear why the relationship between reproductive success and laying date should differ in this manner, although one possibility is that males of the two species defend territories in habits with different seasonal declines in food availability⁶. Further studies are needed to address this question.

Combining the seasonal effect with the reduction in mean offspring reproductive value due to production of hybrid offspring allows us to predict the point at which pairing with a heterospecific male yields the same, or greater, fitness return as pairing with a conspecific male (Fig. 4a). This occurs when hybrid pairing occurs between 3 and 8 days later than the population mean, which will closely approximate the mean breeding date of pure collared pairs, depending on the degree to which reproductive value of hybrid male offspring is discounted relative to pure collareds (see Fig. 3b). Hence, after this point, female collared flycatchers that pair with a heterospecific rather than with a conspecific male should enjoy higher fitness. The relationship between the probability that a female collared flycatcher pairs with a male pied flycatcher and her relative laying date matches this prediction; the probability is at a maximum at +9 days relative to the mean breeding date and greater than that predicted from random mating between +0 and +17 days after the mean breeding date (Fig. 4b). We note that what we have calculated here (Fig. 4a) is the point at which conspecific and heterospecific pairings are of equal value given a simultaneous choice between the two. In fact, the problem for a female collared flycatcher is more likely to be that of choosing whether to accept a male pied flycatcher now, or to reject him and continue searching for a male collared. Thus, the loss in fitness due to late conspecific pairing is likely to be greater than that above.

Discussion

Our findings show that hybrid pairing may sometimes represent adaptive mate choice, because the cost-reducing mechanisms that we identify compensate for the fitness that would otherwise be lost owing to producing unfit hybrids. This has implications for the study of speciation: hybrid zone dynamics may be more complex than previously recognized. Accordingly, predictions from current models of speciation, such as reinforcement²⁸, or sympatric speciation by sexual selection²⁹ may be too simple to capture the full diversity of possible interactions in hybrid zones. Our findings also have implications for understanding the evolution of extra-pair copulation behaviour in birds. Documented benefits for conspecific extra-pair copulation in birds are generally small^{30–32}, and are likely to represent an increase in fitness of only a few per cent³³, as with any

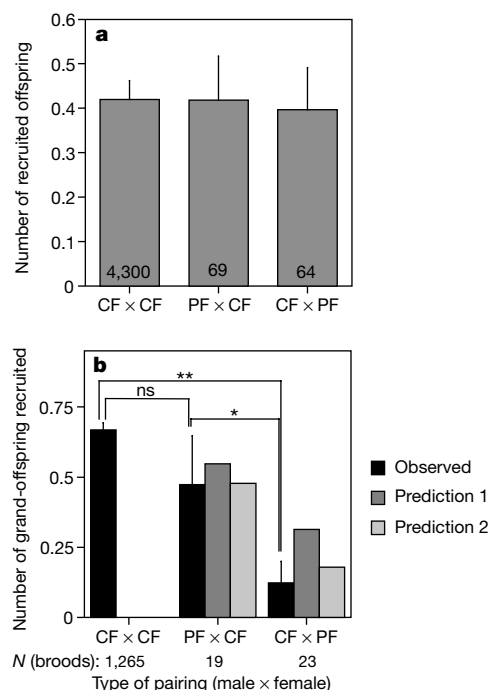


Figure 3 Numbers of recruited offspring and grand-offspring for each pairing type. **a**, Mean ($\pm$ s.e.) reproductive success in relation to pairing type. There is no effect of pairing on reproductive success ($F_{2,4387} = 0.04$, $P = 0.96$). **b**, Number of grand-offspring recruited in relation to pairing type (Kruskal–Wallis test, $H_2 = 9.95$, $P = 0.007$), corresponding to the mean relative reproductive value of offspring produced from the three types of pairings. Values are means ($\pm$ s.e.) for each brood. Pairwise comparisons between the means of each group are shown. Predicted mean reproductive values are shown for two scenarios based on the life-history data in Table 2, and the paternity and sex ratio data in Fig. 1b and Fig. 2a. Prediction 1 assumes that the reduction in fitness of hybrid males is only due to lower fertility, while prediction 2 assumes that the reduction is equivalent to the reduced recruitment success, relative to male collared flycatchers, in Table 2.

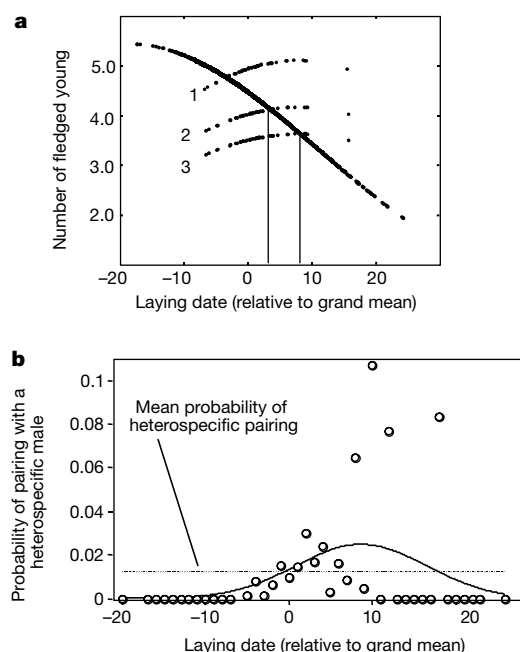


Figure 4 Breeding time, reproductive success, and mating decisions in flycatchers. **a**, Relationship between breeding time (laying date) and number of young fledged from the breeding attempt for pure collared flycatchers and for male pied female collared flycatcher mixed pairs. The lines show the fitted relationships from a general linear model with Poisson error structure, and with terms for laying date ($P < 0.0001$) and laying date² ($P = 0.009$). The relationships differ between the two pair types ($F_{1,3144} = 4.55$, $P = 0.03$). For mixed pairs, three curves are shown: (1) that corresponding to the number of young fledged, ignoring the reduction in mean offspring reproductive value due to production of hybrid offspring; (2) that representing reproductive value discounted by 18% owing to reduced fertility of male and female hybrid offspring (corresponding to prediction 1 in Fig. 3b); (3) that representing reproductive value discounted by 28% owing to reduced recruitment of hybrids (prediction 2 in Fig. 3b). The point at which the curve for the mixed pairs intersects that for the pure pairs is the relative breeding date at which the fitness consequences of accepting a heterospecific mate are equivalent to that of accepting a conspecific mate. **b**, Probability of pairing with a male pied flycatcher in relation to relative breeding date for female collared flycatchers. The points show the proportion of female collared flycatchers laying on a particular date that were paired with a male pied flycatcher. The continuous curve shows the fitted relationship from a logistic regression with date ($P < 0.0001$) and date² ($P = 0.015$); the dotted line shows the mean probability (0.0125) of mixed pairing for the entire data.

form of indirect selection on mating preferences³⁴. In contrast, the benefits from conspecific extra-pair copulation for a female in a mixed pair are large, as they may represent the difference between fertile and sterile offspring. Thus, even if mixed-species pairings are relatively infrequent, selection for extra-pair copulation behaviour may be sufficient to drive this behaviour to fixation. Female birds often prefer extra-pair copulation partners with exaggerated secondary sexual characters^{16,30,35}, and these characters often show the greatest divergence in pairs of sympatric closely related species³⁶. Selection for conspecific extra-pair copulation in sympatry might thus have been an important first step in the evolution of extra-pair copulation behaviour. □

Methods

Life-history data

Data for the Swedish population, on the island of Gotland (57° 10' N, 18° 20' E) were collected as part of a continuous population monitoring study (1980 to present), of flycatchers breeding in artificial nest-boxes. Data on mating frequencies for the Czech population (Dlouhá Loučka: 49° 50' N, 17° 15' E) were collected from 1985 to the present. Laying date, clutch size, and number of young hatching and fledging were recorded by frequent visits to nest-boxes, although data on number of young hatching were not recorded for all breeding attempts. Adults were assigned to breeding attempts on the basis

of their capture (using internal nest-box traps) when feeding young. All adults were ringed with individually numbered aluminium rings on first capture, and all nestlings were ringed before fledging. Reproductive success (number of recruited offspring) was determined from recapture data in subsequent year, and lifetime reproductive success defined as the summed number of recruits over an individual's lifespan. In analyses of long-term data from the Swedish population we excluded all breeding attempts that had been subject to experimental manipulations (such as brood or clutch size manipulation), where these manipulations occurred before the relevant variable was measured. Therefore, the lifetime fitness measures are underestimates, but because experimental treatment was random with respect to species identity, comparisons of their relative sizes are unaffected by this restriction. In addition, for calculating lifetime fledgling production and lifetime reproductive attempts, we excluded all birds that were still alive in 1999, and for calculating lifetime reproductive success all birds that were still alive in 1998 or later, because some offspring do not recruit to the breeding population for two years. The expected relative fitness of the three types of pairing (pure collared; male pied × female collared; and male collared × female pied) was calculated as:

$$1 - (P_f \cdot P_{hs} \cdot (1 - W_f)) - (P_m \cdot P_{hs} \cdot (1 - W_m))$$

where P_f and P_m are the proportion of female and male offspring respectively, and P_{hs} is the proportion of offspring sired by heterospecific males. W_f and W_m are the fitnesses of hybrid females and males respectively, relative to a collared flycatcher offspring of that sex (relative fitness defined as unity). Flycatchers were identified on first capture as either pied, collared or F_1 hybrid based on the size and distribution of white plumage patches (males), the pattern of white on the nape feathers and the plumage tone of the upperparts (females) and on wing length and species-specific calls³⁷. Identification of recruited nestlings was made without knowledge of their parents.

Molecular genetic analyses

Paternity was determined based upon allele-sharing at four polymorphic microsatellite markers: *FhU1*, *FhU2*, *FhU3* and *FhU4*. PCR conditions were as described^{38,39}. The power of these markers to exclude conspecific paternity is not particularly high (approximately 0.96 in both collared flycatchers and pied flycatchers), which suggests that we may have overlooked a few cases of conspecific extra-pair paternity in our analyses of pure (CF × CF and PF × PF) pairs. However, of more relevance to our analyses is the fact that marked species-specific differentiation at these markers³⁹ greatly increases their power to detect cases of conspecific extra-pair paternity in mixed pairs, and heterospecific extra-pair paternity in pure pairs. The sex of nestling flycatchers was determined by PCR amplification of the pair of sex-linked *CHD1* genes, using either SSCP (single-stranded conformation polymorphism) analysis³⁴ or amplification using primers P2 and P8 (ref. 40) followed by silver-staining on 6% polyacrylamide gels. The maternal species identity of F_1 hybrids was determined by amplification of a stretch of mtDNA containing a 32-bp indel, as described¹⁷. Molecular species identification of both extra-pair sires and F_1 hybrids was possible using microsatellites because all four loci used for parentage analysis show some degree of species differentiation in allele size distributions, and in three cases there was little allele-sharing between the two species³⁹. An assignment test⁴¹ based on genotypes at these four loci correctly classified all of 108 adult pied flycatchers and 381 adult collared flycatchers to species. F_1 hybrids were classed as such if they were heterozygous for species-specific alleles at the three loci with largest R_{ST} values³⁹ (R_{ST} is a measure of genetic differentiation at microsatellite loci).

Statistical methods

Data were analysed using general linear modelling techniques⁴² with binomial errors and logit link for proportional data (sex ratio, proportion of nestlings sired by male, hatching and fledging success), with scale correction for over-dispersed data. Count data (number of fledged offspring) were analysed with Poisson errors and a natural logarithm link, again with scale correction for over-dispersion. Changes in deviance were tested against the F -distribution, as recommended when data are overdispersed⁴². Our data contain repeated observation of some males and females; we chose to treat each pairing event as an independent data point since less than 1% of pairings involved the same pair of individuals in different years, and because the within-individual repeatability of life-history traits such as numbers of young fledged or recruited is very low^{13,43}.

Received 10 January; accepted 20 March 2001.

1. Arnold, M. L. *Natural Hybridization and evolution* 7 (Oxford Univ. Press, Oxford, 1997).
2. Grant, P. R. & Grant, B. R. Hybridization of bird species. *Science* **256**, 193–197 (1992).
3. Grant, P. R. & Grant, B. R. Hybridization, sexual imprinting, and mate choice. *Am. Nat.* **149**, 1–28 (1997).
4. Alatalo, R. V., Eriksson, D., Gustafsson, L. & Lundberg, A. Hybridization between pied and collared flycatchers: sexual selection and speciation theory. *J. Evol. Biol.* **3**, 375–389 (1990).
5. Gelter, H. P., Tegelström, H. & Gustafsson, L. Evidence from hatching success and DNA fingerprinting for the fertility of hybrid pied × collared flycatchers. *Ficedula hypoleuca* × *albicollis*. *Ibis* **134**, 62–68 (1992).
6. Saetre, G.-P., Kral, M., Bures, S. & Ims, R. A. Dynamics of a clinal hybrid zone and a comparison with island hybrid zones of flycatchers (*Ficedula hypoleuca* and *F. albicollis*). *J. Zool.* **247**, 53–64 (1999).
7. Saetre, G.-P. et al. A sexually selected character displacement in flycatchers reinforces premating isolation. *Nature* **387**, 589–592 (1997).
8. Lundberg, A. & Alatalo, R. V. *The Pied Flycatcher* 121 (Poyser, London, 1992).
9. Hubbs, C. L. Hybridization between fish species. *System. Zool.* **4**, 1–20 (1955).
10. Saetre, G.-P., Kral, M. & Bures, S. Differential species recognition abilities of males and females in a flycatcher hybrid zone. *J. Avian Biol.* **28**, 259–263 (1997).
11. Real, L. A. Search theory and mate choice. I. Models of single-sex discrimination. *Am. Nat.* **136**, 376–405 (1990).

12. Wiggins, D. A., Pärt, T. & Gustafsson, L. Seasonal decline in collared flycatcher *Ficedula albicollis* reproductive success: an experimental approach. *Oikos* **70**, 359–364 (1994).
13. Merilä, J. & Sheldon, B. C. Lifetime reproductive success and heritability in nature. *Am. Nat.* **155**, 301–310 (2000).
14. Verhulst, S., van Balen, J. H. & Tinbergen, J. M. Seasonal decline in reproductive success in the great tit: variation in time or quality? *Ecology* **76**, 2392–2403 (1995).
15. Daan, S., Dijkstra, C. & Tinbergen, J. M. Family planning in the kestrel (*Falco tinnunculus*): the ultimate control of covariation of laying date and clutch size. *Behaviour* **114**, 83–116 (1990).
16. Sheldon, B. C. & Ellegren, H. Sexual selection resulting from extra-pair paternity in collared flycatchers. *Animal Behav.* **57**, 285–298 (1999).
17. Saetre, G.-P. & Moum, T. A simple molecular method for species identification of pied and collared flycatchers. *Hereditas* **132**, 171–172 (2000).
18. Price, C. S. C. Conspecific sperm precedence in *Drosophila*. *Nature* **388**, 663–666 (1997).
19. Howard, D. J. Conspecific sperm and pollen precedence and speciation. *Annu. Rev. Ecol. Syst.* **30**, 109–132 (1999).
20. Sheldon, B. C., Räsänen, K. & Dias, P. C. Certainty of paternity and paternal effort in the collared flycatcher. *Behav. Ecol.* **8**, 421–428 (1997).
21. Sheldon, B. C. & Ellegren, H. Paternal effort related to experimentally manipulated paternity of male collared flycatchers. *Proc. R. Soc. Lond. B* **265**, 1737–1742 (1998).
22. Gustafsson, L. in *Lifetime Reproduction in Birds* (ed. Newton, I.) 75–88 (Academic, London, 1989).
23. Haldane, J. B. S. Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* **12**, 101–109 (1922).
24. Ellegren, H., Gustafsson, L. & Sheldon, B. C. Sex ratio adjustment in relation to paternal attractiveness in a wild bird population. *Proc. Natl Acad. Sci. USA* **93**, 11723–11728 (1996).
25. Komdeur, J., Daan, S., Tinbergen, J. & Mateman, A. C. Extreme adaptive modification of the sex ratio of the Seychelles warbler's eggs. *Nature* **385**, 522–525 (1997).
26. Nager, R. G., Monaghan, P., Griffiths, R., Houston, D. C. & Dawson, R. Experimental demonstration that offspring sex ratio varies with maternal condition. *Proc. Natl Acad. Sci. USA* **96**, 570–573 (1999).
27. Sheldon, B. C., Andersson, S., Griffith, S., Örnborg, J., Sendecka, J. Ultraviolet colour variation influences blue tit sex ratios. *Nature* **402**, 874–877 (1999).
28. Liou, L. W. & Price, T. D. Speciation by reinforcement of pre-mating isolation. *Evolution* **48**, 1451–1459 (1994).
29. Higashi, M., Takimoto, G. & Yamamura, N. Sympatric speciation by sexual selection. *Nature* **402**, 523–526.
30. Hasselquist, D., Bensch, S. & von Schantz, T. Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler. *Nature* **381**, 229–232 (1996).
31. Sheldon, B. C., Merilä, J., Qvarnström, A., Gustafsson, L. & Ellegren, H. Paternal genetic contribution to offspring condition predicted by relative size of male secondary sexual character. *Proc. R. Soc. Lond. B* **264**, 297–302 (1997).
32. Johnsen, A., Anderson, V., Sunding, C. & Lifjeld, J. T. Female bluethroats enhance offspring immunocompetence through extra-pair copulations. *Nature* **406**, 296–299 (2000).
33. Møller, A. P. & Alatalo, R. V. Good genes effects in sexual selection. *Proc. R. Soc. Lond. B* **266**, 85–91 (1999).
34. Kirkpatrick, M. & Barton, N. H. The strength of indirect selection on female mating preferences. *Proc. Natl Acad. Sci. USA* **94**, 1282–1286 (1997).
35. Møller, A. P. & Ninni, P. Sperm competition and sexual selection: a meta-analysis of paternity studies of birds. *Behav. Ecol. Sociobiol.* **43**, 345–358 (1998).
36. Price, T. Sexual selection and natural selection in bird speciation. *Phil. Trans. R. Soc. Lond. B* **353**, 251–260 (1998).
37. Mild, K. & Shiriha, H. Field identification of pied, collared and semi-collared flycatchers. Part 2: females in breeding plumage. *Bird. World* **7**, 231–240 (1994).
38. Sheldon, B. C. & Ellegren, H. Offspring sex and paternity in the collared flycatcher. *Proc. R. Soc. Lond. B* **263**, 1017–1021 (1996).
39. Saetre, G.-P. *et al.* Speciation, introgressive hybridization and non-linear rate of molecular evolution in flycatchers. *Mol. Ecol.* **10**, 737–749 (2001).
40. Griffiths, R., Double, M. C., Orr, K. & Dawson, R. J. G. A DNA test to sex most birds. *Mol. Ecol.* **7**, 1071–1075 (1998).
41. Paetkau, D., Calvert, W., Stirling, I. & Strobeck, C. Microsatellite analysis of population structure in Canadian polar bears. *Mol. Ecol.* **4**, 347–354 (1995).
42. Crawley, M. J. *GLIM for Ecologists* 279 (Blackwell, Oxford, 1993).
43. Przybylo, R., Sheldon, B. C. & Merilä, J. Climatic influences on breeding and morphology evidence for phenotypic plasticity. *J. Anim. Ecol.* **69**, 395–403 (2000).

Acknowledgements

We thank the many people that helped to collect data from the Swedish and Czech study populations, particularly J. Träff for providing access to mixed pairs in Sweden, and M. Kral for access to Czech data. P. R. Grant, L. E. B. Kruuk, J. Merilä and T. Price commented on a previous version of the manuscript. The work was funded by grants from the Swedish Natural Sciences Research Council, the Norwegian Research Council, an Erasmus grant to T.V., and by a Royal Society University Research Fellowship to B.C.S.

Correspondence and requests for materials should be addressed to B.C.S. (e-mail: ben.sheldon@zoology.oxford.ac.uk).

Four-terminal resistance of a ballistic quantum wire

R. de Picciotto*, H. L. Stormer*†, L. N. Pfeiffer*, K. W. Baldwin* & K. W. West*

* Bell-Labs, Lucent Technologies, Murray Hill, New Jersey 07974, USA

† Departments of Physics and Applied Physics, Columbia University, New York 10003, USA

The electrical resistance of a conductor is intimately related to the relaxation of the momentum of charge carriers. In a simple model, the accelerating force exerted on electrons by an applied electric field is balanced by a frictional force arising from their frequent collisions with obstacles such as impurities, grain boundaries or other deviations from a perfect crystalline order¹. Thus, in the absence of any scattering, the electrical resistance should vanish altogether. Here, we observe such vanishing four-terminal resistance in a single-mode ballistic quantum wire. This result contrasts the value of the standard two-probe resistance measurements of $h/2e^2 \approx 13 \text{ k}\Omega$. The measurements are conducted in the highly controlled geometry afforded by epitaxial growth onto the cleaved edge of a high-quality GaAs/AlGaAs heterostructure. Two weakly invasive voltage probes are attached to the central section of a ballistic quantum wire to measure the inherent resistance of this clean one-dimensional conductor.

Electronic transport with no scattering can occur in nanoscale solid-state devices^{2–8}. Conceptually, the simplest of these structures is the ballistic one-dimensional wire^{6–8}, in which the transverse motion is quantized into discrete modes, and the longitudinal motion is free. In this case electrons are envisioned to propagate freely down a clean narrow pipe. However, the actual resistance of such a wire is found to be very different from zero. Instead, its value is the resistance quantum ($R_0 = h/2e^2$) divided by the number of occupied transverse modes⁹. Hence, in the quantum limit, when the pipe is sufficiently narrow to support only a single mode, the resistance of a perfect wire is rather large: $R_0 \approx 13 \text{ k}\Omega$.

The origin of this resistance is best appreciated in the framework of a model for one-dimensional conduction proposed by Landauer^{10,11}. We consider two electron reservoirs connected by a perfect single-mode wire. To maintain a current, I , through the wire, a higher electrochemical potential is imposed on the right-hand side (r.h.s.) reservoir, μ_r^{in} , than on the left-hand side (l.h.s.) reservoir, μ_l^{in} . Electrons that propagate through the wire to the left, away from the r.h.s. reservoir, maintain its electrochemical potential, $\mu_l = \mu_r^{\text{in}}$. Conversely, electrons that propagate to the right maintain the electrochemical potential of the l.h.s. reservoir, $\mu_r = \mu_l^{\text{in}}$ (see Fig. 1c). In the absence of scattering, these electrochemical potentials do not vary along the wire.

However, although both μ_l and μ_r are uniform throughout the length of the wire, both vary at its ends, where the wire is connected to the reservoirs. These steps arise from electron scattering that equilibrates left movers and right movers with the local electrochemical potential of the reservoir. The variations in μ_l and μ_r can be viewed as contact resistances to the wire^{12,13}, each having a value $R_c = R_0/2$. This is the origin of the relatively high resistance observed in two-terminal measurements on a quantum wire. On the other hand, within the wire the electrochemical potentials are independent of position and the result of a four-terminal resistance measurement should be zero, in the ideal case. We will refer to this as vanishing 'intrinsic resistance'.

Here we report the realization of a four-terminal geometry, which allows us to measure this intrinsic resistance of a quantum wire. The wires are fabricated from GaAs/AlGaAs heterostructures using the cleaved-edge overgrowth (CEO) technique¹⁴ as illustrated in Fig. 1a.

The resultant wire resides all along an atomically precise edge of a GaAs quantum well. The quantum well itself supports a two-dimensional electron gas (2DEG) that is coupled to the wire from the side. We use prefabricated top gate electrodes (see Fig. 1a) to shape this 2DEG sheet. Biasing any one of the gates depletes the 2DEG underneath and creates a stretch of one-dimensional wire in front of it, which is now separated from the 2DEG. The width of the gate defines the length of this isolated wire section, and the 2DEG areas on either side conveniently serve as source and drain contacts to the wire. This geometry lends itself to a straightforward

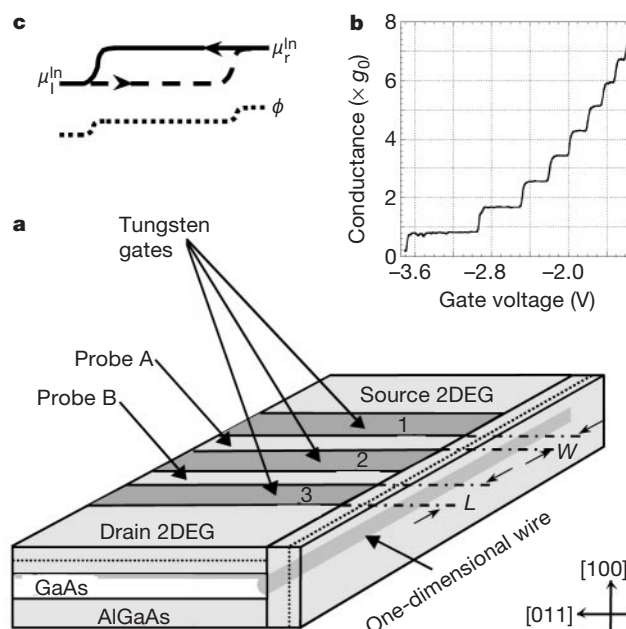


Figure 1 Electronic transport in cleaved-edge overgrowth quantum wires. **a**, Geometry of the CEO device. The fabrication starts with a high-quality 2DEG created by epitaxial growth of a unilaterally doped GaAs quantum well onto a [001] GaAs substrate. The resultant 2DEG has a carrier density $n_s \approx 2.5 \times 10^{11} \text{ cm}^{-2}$, and mobility $\mu \approx 4 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. Subsequently, this wafer is cleaved inside the MBE chamber to expose a clean and atomically smooth [110] surface, which is immediately overgrown with a modulation-doped epitaxial-layer sequence. The additional remote Si dopants that are introduced by this overgrowth step lead to a higher electron density near the cleaved edge of the quantum well. As in conventional modulation-doped samples, a strong built-in electric field binds this excess charge to the cleaved-edge interface, creating one-dimensional bound states all along the edge of the GaAs quantum well. This wire contains about 10 electronic modes and coexists with a 2DEG that resides in the quantum-well plane and couples to the wire from the side. To separate the 2DEG from the wire, prefabricated tungsten gate electrodes (for example, gate 1) are used. They deplete the 2DEG underneath them but preserve the one-dimensional channel in this region along the edge. The width of the tungsten gate defines the length, L , of the isolated wire section. Increasing the gate voltage beyond depletion of the 2DEG provides a convenient tool to control the number of occupied one-dimensional modes in this wire section. Three gates allow us to separate out four distinct 2DEG regions (see text). **b**, A representative result of a two-terminal conductance measurement of a CEO wire at a temperature, $\theta = 300 \text{ mK}$. For this measurement, only one electrode (say, number 1) is activated, which creates in its front a $2\text{-}\mu\text{m}$ -long section of an isolated one-dimensional wire. Clear conductance plateaux arise, attesting to the high quality of the wire. Practically identical behaviour is observed for the wire sections in front of all electrodes (not shown). The value of the quantized resistance is somewhat larger than the universal value of $R_0^{7,8,19}$. The origin of this deviation is non-ideal coupling between the CEO wire and its 2DEG source and drain contacts¹⁹. Because the goal of the present paper is the determination of the intrinsic wire resistance by a four-probe measurement, which is insensitive to any contact resistance, this non-ideal current injection is inconsequential. **c**, The spatial behaviour of the electrochemical potentials of left and right movers, μ_l and μ_r respectively, and of the electrical potential, ϕ , in a two-terminal geometry.

two-probe measurement on a quantum wire. Figure 1b shows the high-quality, quantized conductance steps that are observed in such a specimen as successive one-dimensional sub-bands are depopulated.

To create the more complex contacting geometry for a four-probe measurement, we use a pattern of three successive electrodes (numbers 1–3 in Fig. 1a). This configuration establishes two narrow 2DEG strips (A and B), that serve as voltage taps into the central portion of the wire. The wide 2DEG regions to the far right and far left serve as source and drain contacts for the current, as in the two-probe case. This gate configuration offers unique flexibility. Activating only electrode 2 provides a two-terminal measurement on the central section of the wire. Biasing all three electrodes creates an extended quantum wire along the edge and separates voltage probes A and B from the source and drain respectively. This allows us to perform a four-probe measurement on the same central section of the wire. Such a small change in the biasing scheme of our CEO sample switches directly from two-probe to four-probe geometry with marked consequences.

Figure 2 shows the result of a two-probe and a four-probe measurement on the same CEO quantum wire. The width of each

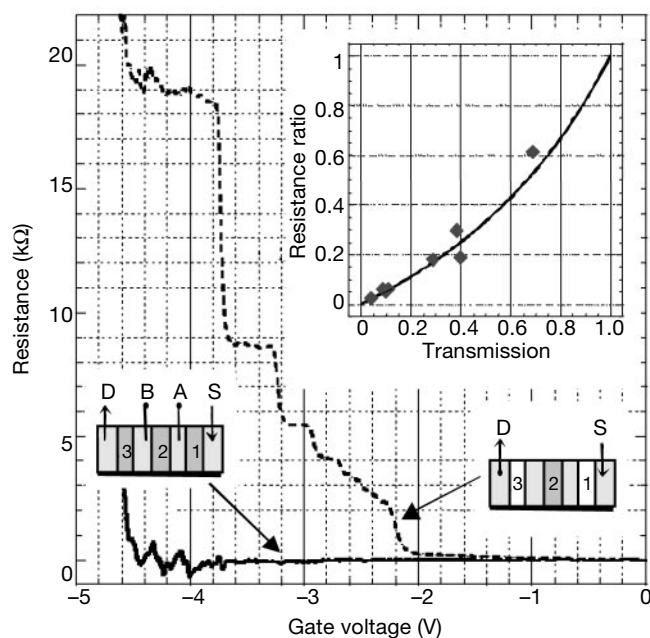


Figure 2 Two- and four-terminal resistances of a ballistic quantum wire. The dashed line shows the two-terminal resistance of the 2- μm -long central section of the wire versus the voltage applied to the associated gate 2. Gates 1 and 3 are not activated. The solid line shows the four-terminal resistance, $(V_A - V_B)/I$, versus the voltage applied to gate 2. Here V_A and V_B are the voltages at probes A and B respectively and I is the current driven from source to drain. For this measurement, the voltages applied to gates 1 and 3 correspond to a single mode in the wire sections in front of these gates. Measurements were performed at a temperature of $\theta = 300$ mK with an excitation current smaller than 1 nA. While the two-terminal resistance moves through the characteristic quantized resistance steps, the four-terminal resistance fluctuates around zero indicating that the inherent resistance of a clean one-dimensional wire is vanishingly small. The small oscillations around zero resistance (from -3.8 V to -4.5 V) suggest that mesoscopic variation of the various transmission amplitudes with the one-dimensional density dominate the resistance in this regime. Indeed a similar, although not identical, pattern is observed upon successive cool-downs of the same device. As expected, similar mesoscopic variations are observed when a magnetic field is applied (see Fig. 3). Inset, probe invasiveness in a quantum wire. Diamonds, the ratio between the four-terminal and two-terminal resistances versus the invasiveness of the voltage probes (see text). Solid line, theoretical prediction of the Landauer–Buttiker model¹⁶ (see text). All measurements are for single-mode wires.

gate electrode is $L = 2 \mu\text{m}$. The gates are separated by 2DEG strips of width $W = 2 \mu\text{m}$ each. For the four-probe measurement electrodes 1 and 3 are set to maintain a single transverse mode in the respective wire sections, whereas these gates are not activated in the two-probe case. In both measurements the bias to the central electrode is scanned, depopulating successive one-dimensional sub-bands. As is evident from Fig. 2, the two-terminal resistance moves through the characteristic quantized resistance steps, whereas the four-terminal resistance hovers around zero resistance, with small mesoscopic fluctuations superimposed. This is the central result of our experimental work: the inherent resistance of a clean one-dimensional wire is vanishingly small.

It demonstrates experimentally that the specific two-terminal resistance value stems from an existing ‘contact resistance’ between the wire and the two macroscopic electron reservoirs at either end. In a two-terminal resistance measurement this inherent contact resistance adds to the intrinsic resistance of the wire and hence, even for a perfect wire, the two-terminal resistance is at least R_0 per mode¹³. We will now discuss in detail the physics of two-probe versus four-probe measurements on a quantum wire.

A four-terminal geometry is a common method used to circumvent contributions from the contacts to the measured resistance. Typically, a current is driven by two contacts at the far ends of a long rectangular specimen. The electrochemical potential drop along the current path is determined by two separate voltage probes, located near the centre of the specimen, far away from the current contacts. To deduce the intrinsic resistivity, it is essential that the voltage probes do not disturb the current flow. In two or three dimensions, this is readily accomplished by using small voltage probes that leave contiguous regions of the specimen intact. In one dimension this cannot be accomplished for obvious geometrical reasons and voltage probes are always ‘invasive’. This invasiveness of a voltage probe is intimately related to the probability that an electron passing the probe will scatter into it^{15–19}. In general, the transmission probability of left movers, T_l , and right movers, T_r , may differ.

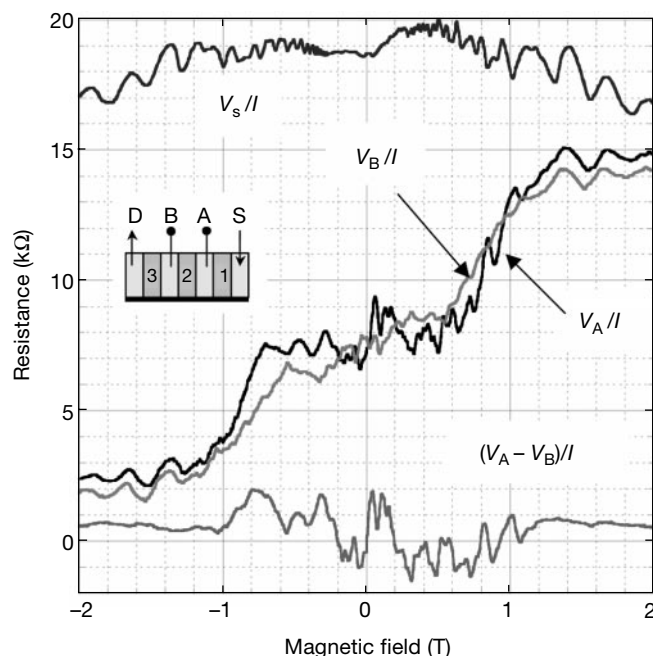


Figure 3 Magnetic field dependence of probe coupling. Magnetic field dependence of the two-terminal (V_S/I), three-terminal (V_B/I and V_A/I) and four-terminal ($(V_A - V_B)/I$) resistances. The sections of the wire in front of all three gates support a single mode. The invasiveness of both probes is about 4%. The voltages at both taps vary from $V \approx V_S$ (voltage at source) at large positive field to $V \approx V_D = 0$ (voltage at drain) at large negative field.

Yet, in the simplest case of symmetric coupling, the invasiveness is proportional to $T = T_1 = T_r = \frac{1}{2}(R_{2r}/R)$ with R_{2r} and R being the two-terminal resistance and the resistance of the probe–wire interface respectively. The smaller T is, the less invasive is the voltage probe, leaving the wire practically undisturbed^{17,19}.

The results shown in Fig. 2 correspond to a device with two 2- μm -wide 2DEG probes. The resistance between each probe and wire is about 250 k Ω (not shown) indicating that their invasiveness is merely about 4%. Moreover, the overall source–drain resistance of the three wire sections in front of the electrodes, each having one occupied transverse mode, is about 20 k Ω (not shown). This value is only slightly larger than the 19 Ω two-terminal resistance of the 2- μm -long central section (see Fig. 2), whereas classically we would have expected at least a tripling of the resistance. The absence of a significant increase of the overall resistance further indicates the low invasiveness of the voltage probes and the negligible back scattering in the wire.

Our device geometry allows extensive control over the 2DEG–wire coupling and hence over the invasiveness of a contact to a CEO wire. For example, the gap between electrodes 1 and 2 controls the invasiveness of probe A. It can be made weaker (stronger) by reducing (increasing) the width, W , as compared to the two- to one-dimensional scattering length. This scattering length was previously determined to be approximately 6 μm in our wires¹⁹, and is easily achieved with electrode separation established by standard photolithography.

The inset of Fig. 2 shows a plot of the ratio, α , of the four-terminal to the two-terminal resistance, against the transmission, T . We find that α increases monotonically with T , approaching unity together with T itself. These data are in good agreement with theoretical predictions¹⁶ for a four-terminal measurement with symmetrically coupled probes of invasiveness T : $\alpha = T/(2 - T)$. It clearly demonstrates the crucial role played by the invasiveness of the voltage probes in a four-terminal resistance measurement of a ballistic quantum wire. In the limit of fully invasive probes, $T \rightarrow 1$, the voltage probes essentially break the wire into three separate pieces and thus defeat their purpose, because in this limit a four-terminal measurement yields the same value as a two-terminal one. These results demonstrate the special role played by quantum coherence in the series addition of one-dimensional wires. A very invasive voltage probe acts as a strong inelastic scatterer that collapses the gap between the electrochemical potentials of left and right movers¹⁷. Such a collapse at the location of each probe leads to the well known addition rule for classical resistors.

In a ballistic one-dimensional wire, left and right propagating electrons acquire two distinct electrochemical potentials. This behaviour is unique to one dimension and is a direct result of the absence of back scattering. The idea of a local electrochemical potential, and therefore of the meaning of a voltage measurement, is thus exceptional in a ballistic wire. A voltage measurement corresponds to the value of the probe's electrochemical potential required to null the probe–sample current. In diffusive samples the probe acquires the local electrochemical potential of the neighbouring sample region. In a ballistic wire, however, the electrochemical potential required to null the probe–wire current is, in general, $\mu_p = \alpha\mu_l + \beta\mu_r$, with $\alpha = T_l/(T_r + T_l)$ and $\beta = T_r/(T_r + T_l)$. For example, a probe that couples only to left movers ($T_l = 1$ and $T_r = 0$) will measure μ_l while a probe that couples only to right movers will measure μ_r . A ballistic wire has two distinct electrochemical potentials and thus there are many ways to define the internal resistance.

This ambiguity can be resolved by requiring charge neutrality. The spatial charge density of electrons, when averaged over a length scale larger than the screening length, is always compensated by a fixed positive background charge, leaving the sample as a whole neutral. In two or three dimensions, this implies $\mu(x) - \phi(x) = C$, with ϕ the local potential and C a translation invariant constant.

Thus, the electrochemical potential difference in a four-terminal configuration yields the potential drop between voltage probes. On the other hand, in one dimension charge neutrality implies $\mu(x) - \phi(x) = C$, with $\bar{\mu} = (\mu_l + \mu_r)/2$ (see Fig. 1c). Therefore, only a symmetric voltage probe, with $T_r = T_l$ and thus $\mu_p = \bar{\mu}$, can measure the local one-dimensional potential. This analogy between the one-dimensional case and the two- to three-dimensional case makes it seem natural¹⁵ to define the intrinsic resistance of a ballistic wire in terms of $\bar{\mu}$, rather than in terms of μ_l or μ_r .

The voltage probes of our four-terminal geometry are highly symmetric. To illustrate this, we measure the voltage of each probe, V_A and V_B , with respect to the drain. We find both voltages to be about half the voltage applied from source to drain, although the distance from the drain to probe A is twice the distance to probe B. This result shows that both probes measure $\bar{\mu}$ regardless of their location and also confirms the view of two equal 'contact resistances' existing at each end of the wire. To further illustrate the effect of the probe–wire coupling symmetry, we measure V_A and V_B as a function of a magnetic field applied perpendicular to the quantum-well plane, as shown in Fig. 3. The effect on the 2DEG is small since at 1 T the filling factor is $\nu \approx 10$. With increasing positive magnetic field, the voltage at both probes increases, and saturates at approximately 1 T with a value close to the source voltage. Conversely, a negative magnetic field reduces the voltages to a value close to the voltage of the drain. This voltage swing originates from the Lorentz force exerted on the carriers. A positive (negative) magnetic field forces the left-moving (right-moving) electrons in the wire closer to the 2DEG probes. Hence the magnetic field breaks the symmetry of the coupling by establishing preferential coupling between the probes and either left or right movers, depending on the field direction. When only left (right) movers couple to the probes their voltages approach the source (drain) voltage. To fully break this symmetry requires a magnetic field of about 1 T, which corresponds to a magnetic length of approximately 250 Å, in good agreement with the width of our wire of about 200 Å.

Thus we were able to perform four-terminal resistance measurements on ballistic quantum one-dimensional wires and observed a vanishing resistance. This demonstrates experimentally that the high resistance values observed in two-probe measurements on such systems originate from the contacts alone, and that the intrinsic resistance is negligible. Our results validate the view of the quantized resistance being twice the minimal contact resistance to a single wire mode. □

Received 3 January; accepted 4 March 2001.

- Drude, P. Zur elektronentheorie der metalle 1 Teil. *Ann. Phys.* **1**, 566–613 (1900); Zur elektronentheorie der metalle 2 Teil, Galvanomagnetische und thermomagnetische effecte. *Ann. Phys.* **3**, 369–402 (1900).
- Heiblum, M. Tunneling hot electron transfer amplifiers (THETA): Amplifiers operating up to the infrared. *Solid State Electron.* **24**, 343–366 (1981).
- van Wees, B. J. et al. Quantized conductance of point contacts in a two-dimensional electron gas. *Phys. Rev. Lett.* **60**, 848–850 (1988).
- Wharam, D. A. et al. One-dimensional transport and the quantisation of the ballistic resistance. *J. Phys. C* **21**, L209–L214 (1988).
- van Houten, H. et al. Coherent electron focusing with quantum point contacts in a two-dimensional electron gas. *Phys. Rev. B* **39**, 8556–8575 (1989).
- Tarucha, S., Honda, T. & Saku, T. Reduction of quantized conductance at low temperature observed in 2 to 10 μm long quantum wires. *Solid State Commun.* **94**, 413–418 (1995).
- Yacoby, A. et al. Nonuniversal conductance quantization in quantum wires. *Phys. Rev. Lett.* **77**, 4612–4615 (1996).
- Rother, M., Wegscheider, W., Bichler, M. & Abstreiter, G. in *Proc. of ICPS24 Vol. 2* (ed. Gershoni, D.) Th-P137 (World Scientific, Singapore, 1998).
- Imry, Y. in *Introduction to Mesoscopic Physics* (eds Craighead, H. G. et al.) 89–120 (Oxford Univ. Press, Oxford, 1997).
- Landauer, R. Spatial variations of currents and fields due to localized scatterers in metallic conduction. *IBM J. Res. Dev.* **1**, 223–231 (1957).
- Landauer, R. Electrical resistance of disordered one-dimensional lattices. *Phil. Mag.* **21**, 863–867 (1970).
- Imry, Y. in *Directions in Condensed Matter Physics* (eds Grinstein, G. & Mazenko, G.) (World Scientific, Singapore, 1986).
- Glazman, L. I., Lesovik, G. B. D., Khmelnitskii, E. & Shekhter, R. I. Reflectionless quantum transport and fundamental ballistic-resistance steps in microscopic constrictions. *JEPT Lett.* **48**, 238–241 (1988).
- Pfeiffer, L. N. et al. Transport and optics in quantum wires fabricated by MBE overgrowth on the (110) cleaved edge. *Microelectron. J.* **28**, 817–823 (1997).

15. Engquist, H. L. & Anderson, P. W. Definition and measurement of the electrical and thermal resistances. *Phys. Rev. B* **24**, 1151–1154 (1981).
16. Buttiker, M. Four-terminal phase-coherent conductance. *Phys. Rev. Lett.* **57**, 1761–1764 (1986).
17. Buttiker, M. Role of quantum coherence in series resistors. *Phys. Rev. B* **33**, 3020–3026 (1986).
18. Buttiker, M. Symmetry of electrical conduction. *IBM J. Res. Dev.* **32**, 317–323 (1988).
19. de Picciotto, R. *et al.* 2D–1D coupling in cleaved edge overgrowth. *Phys. Rev. Lett.* **85**, 1730–1733 (2000).

Acknowledgements

We thank A. L. Moustakas, S. H. Simon, A. Yacoby and C. M. Varma for fruitful discussions.

Correspondence and requests for materials should be addressed to R.d.P. (e-mail: rd25@lucent.com).

Superconductivity in the non-oxide perovskite MgCNi_3

T. He*, Q. Huang†, A. P. Ramirez‡, Y. Wang§, K. A. Regan*, N. Rogado*, M. A. Hayward*, M. K. Haas*, J. S. Slusky*, K. Inumara*, H. W. Zandbergen*, N. P. Ong§ & R. J. Cava*

* Department of Chemistry and Princeton Materials Institute; § Department of Physics, Princeton University, Princeton, New Jersey, USA

† Department of Materials and Nuclear Engineering, University of Maryland, College Park, Maryland; and NIST Center for Neutron Research, Gaithersburg, Maryland, USA

‡ Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, Los Alamos, New Mexico, USA

The interplay of magnetic interactions, the dimensionality of the crystal structure and electronic correlations in producing superconductivity is one of the dominant themes in the study of the electronic properties of complex materials. Although magnetic interactions and two-dimensional structures were long thought to be detrimental to the formation of a superconducting state, they are actually common features of both the high transition-temperature (T_c) copper oxides and low- T_c material Sr_2RuO_4 , where they appear to be essential contributors to the exotic electronic states of these materials¹. Here we report that the perovskite-structured compound MgCNi_3 is superconducting with a critical temperature of 8 K. This material is the three-dimensional analogue of the $\text{LnNi}_2\text{B}_2\text{C}$ family of superconductors, which have critical temperatures up to 16 K (ref. 2). The itinerant electrons in both families of materials arise from the partial filling of the nickel d -states, which generally leads to ferromagnetism as is the case in metallic Ni. The high relative proportion of Ni in MgCNi_3 suggests that magnetic interactions are important, and the lower T_c of this three-dimensional compound—when compared to the $\text{LnNi}_2\text{B}_2\text{C}$ family—contrasts with conventional ideas regarding the origins of superconductivity.

The variable stoichiometry compound MgC_xNi_3 (where $0.5 > x > 1.25$) has been previously reported; it was supposed to have a perovskite structure by analogy^{3,4}. Neither its crystal structure nor its physical properties had been determined previously. In this study, samples with nominal formula MgC_xNi_3 for $x = 1.5, 1.25, 1.1, 1.0$ and 0.9 were prepared. The starting materials were bright Mg flakes (Aldrich Chemical), fine Ni powder (99.9% Johnson Matthey), and glassy carbon spherical powder (Alfa AESAR). The starting materials were mixed in 0.5-g batches, and pressed into pellets. The pellets were placed on Ta foil, which was, in turn, placed on an Al_2O_3 boat, and fired in a quartz tube furnace under a mixed gas of 95% Ar and 5% H_2 . The samples were heated for half an hour at 600 °C, followed by one hour at 900 °C. After cooling, they were

ground, pressed into pellets, and heated for an additional hour at 900 °C. Owing to the volatility of Mg encountered during the synthesis of this compound, 20% Mg in excess of the stoichiometric ratio was employed in the initial mixtures. The structural determination, described below, indicated that the compound formed was stoichiometric in metals and that no excess Mg remained in the samples.

The crystal structure of a superconducting sample ($T_c \equiv 7.3$ K, determined magnetically) of nominal composition $\text{MgC}_{1.25}\text{Ni}_3$ was determined by powder neutron diffraction (Fig. 1). The formula for the superconducting phase was found to be $\text{MgC}_{0.96}\text{Ni}_3$. The compound has the classical cubic perovskite structure, space group $\text{Pm}\bar{3}\text{m}$, with lattice parameter $a = 3.81221(5)$ Å. The positions for the atoms are: Mg 1a (0,0,0); C 1b (0.5,0.5,0.5); and Ni 3c (0,0.5,0.5). The temperature factors are 0.90(3), 0.54(4), and 0.75(1) Å² for Mg, C and Ni, respectively. Refinements were performed with variable stoichiometry allowed for the C site. The C site occupancy was found to be 0.960(8), making the exact stoichiometry $\text{MgC}_{0.96}\text{Ni}_3$. In agreement with what is expected from the nominal composition, a small amount of unreacted graphite (2 wt%) was found in the sample. The sensitivity of the superconductivity to the C content of the perovskite phase made it necessary for carbon excess to be added to the initial mixtures to ensure attainment of the superconducting composition. The perovskite crystal structure for MgCNi_3 is shown in the inset to Fig. 1. Comparison to a familiar oxide perovskite such as CaTiO_3 , for example, indicates the structural equivalencies between Ca and Mg, Ti and C, and O and Ni.

The magnetic characterization of the superconducting transitions is shown in Fig. 2. The magnetic onset for the superconducting transition ranges between 7.1 K for a nominal carbon content of 1.1 per MgC_xNi_3 to 7.4 K for nominal carbon content $x = 1.5$. The superconducting transition turns off abruptly for nominal C contents between $x = 1.1$ and $x = 1.0$. All samples in the range of carbon contents shown appear to be single phase by powder X-ray diffraction (which would not be sensitive to the presence of graphite

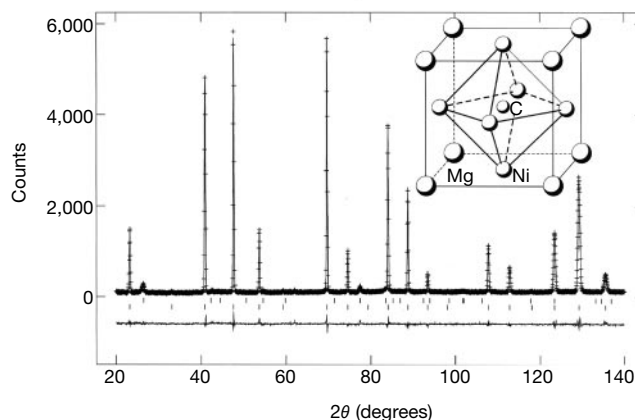


Figure 1 The powder neutron diffraction pattern at ambient temperature for the sample of nominal composition $\text{MgC}_{1.25}\text{Ni}_3$ and the perovskite crystal structure for the superconducting compound MgCNi_3 (inset). Neutrons of wavelength 1.5402 Å were employed (Cu 311 monochromator), with collimators of 15', 20' and 7' of arc before and after the monochromator, and after the sample, respectively. The neutron scattering lengths employed in the structure refinement were 0.538, 0.665 and 1.030 (cm⁻¹) for Mg, C and Ni, respectively. Data are shown as crosses, and the difference plot between model and data shown directly below. The vertical lines (bottom) show the Bragg peak positions for the MgCNi_3 phase. The sample contains 2 wt% graphite (about 25 mol.%) in agreement with the nominal composition. Positions of the graphite peaks are shown as vertical lines above those for MgCNi_3 . The refinement agreement, weighted profile agreement, and χ^2 values obtained were $R = 5.14\%$, $R_w = 6.39\%$ and $\chi^2 = 1.258$, indicating the high quality of the structural model.

impurity). Their crystallographic cell parameters vary smoothly between 3.805(1) Å at nominal $x = 1.0$ and 3.816(1) Å at nominal $x = 1.5$ owing to a difference in carbon content of the perovskite phase. The neutron diffraction results indicate that the composition of the superconducting phase for nominal $x = 1.25$ is $\text{MgC}_{0.96}\text{Ni}_3$. Therefore the data suggest that the highest T_c in the perovskite is obtained for the perfectly stoichiometric composition $\text{MgC}_{1.0}\text{Ni}_3$, which requires the presence of excess carbon to form under our synthetic conditions. Further, the data indicate that the superconductivity disappears abruptly for carbon contents between 0.96 and approximately 0.90 per formula unit. The reason for this is not yet known.

The temperature-dependent resistivity for MgCNi_3 between 290 and 5 K (sample of nominal composition $\text{MgC}_{1.5}\text{Ni}_3$) is shown in Fig. 3. The normal-state resistivity at room temperature is relatively low, even for the polycrystalline sample, about $90 \mu\Omega \text{ cm}$. The resistivity ratio, $\rho_{300\text{K}}/\rho_{9\text{K}}$, is approximately 2.1. The temperature dependence is consistent with what is expected for a poor metal, but whether the shape of $\rho(T)$ is intrinsic or not remains to be verified by future work on single crystals if they become available. The superconducting transition (inset Fig. 3) is very sharp, indicating that the fraction of material with the optimum carbon stoichiometry for the highest T_c is above the percolation threshold. The midpoint of the resistive transition is 8.4 K, the 90%–10% transition width is 0.1 K, and the resistive onset temperature is 8.5 K.

The characterization of the superconducting transition by specific heat measurements is shown in Fig. 4. Figure 4a shows the specific heat divided by temperature ($C(T)/T$) versus T . This figure allows us to perform an equal-area extrapolation to determine the midpoint of the thermodynamic transition, as well as the jump, $\Delta C/T_c$, assuming a mean-field transition. The transition midpoint is 6.2 K, significantly below the resistive onset. This value, as well as the breadth of the jump, ~ 1.7 K, is probably due to a small carbon stoichiometry variation in the sample: the transition width implies that the sample studied possessed a variation of ~ 0.01 in C content. We find $\Delta C/T_c = 19 \text{ mJ per mol Ni K}^{-2}$.

In Fig. 4b we plot $C(T)/T$ versus T^2 . We assume that $C(T)$ has two leading contributions: a linear Fermi-liquid term, γT , where γ is the Sommerfeld constant, and a cubic lattice term. This plot then allows an extrapolation of the normal-state behaviour to $T = 0$ where γ is determined as the $T = 0$ value of $C(T)/T$. By this method, we find that $\gamma \approx 10 \text{ mJ per mol Ni K}^{-2}$ (Both γ and $\Delta C/T_c$ for MgCNi_3 are comparable to the values found for $\text{LuNi}_2\text{B}_2\text{C}$, 9 and $20 \text{ mJ per mol Ni K}^{-2}$, respectively⁵.) If we assume that superconductivity is

mediated by phonons, then $\Delta C/T_c$ is related to γ by a numerical solution of the Eliashberg equations^{6,7}:

$$\Delta C/T_c = (1.43 + 0.042\lambda_{\text{ph}}^2 - 0.195\lambda_{\text{ph}}^3)\gamma$$

where λ_{ph} is the electron–phonon coupling constant. From the above data we ascertain $\Delta C/T_c/\gamma = 1.9 \pm 0.1$, which leads to $\lambda_{\text{ph}} = 0.77^{+0.17}_{-0.09}$. This value of λ_{ph} is in the range of conventional phonon-mediated pairing, though given the large amount of Ni in the compound, it is too early to attribute the superconductivity to a conventional microscopic mechanism.

Although the structural analogy between superconducting intermetallic perovskite MgCNi_3 and superconducting oxide perovskites

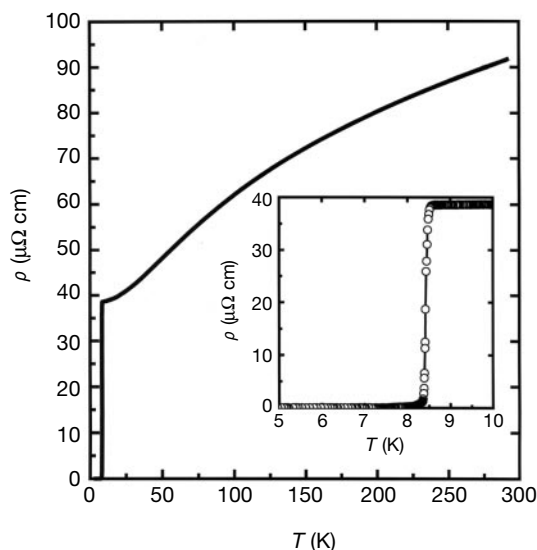


Figure 3 The temperature-dependent resistivity for MgCNi_3 between 290 and 5 K (sample of nominal composition $\text{MgC}_{1.5}\text{Ni}_3$). We measured the resistivity of a polycrystalline pellet made from powder that was first reacted as described, and then pressed at a pressure of 10 kbar for one hour at 700 °C. The four-probe a.c. method was used with current 100 mA (for the normal state), 10 mA (for the transition region), and $f = 16.0 \text{ Hz}$. The normal-state resistivity and a detailed view (inset) of the superconducting transition are shown.

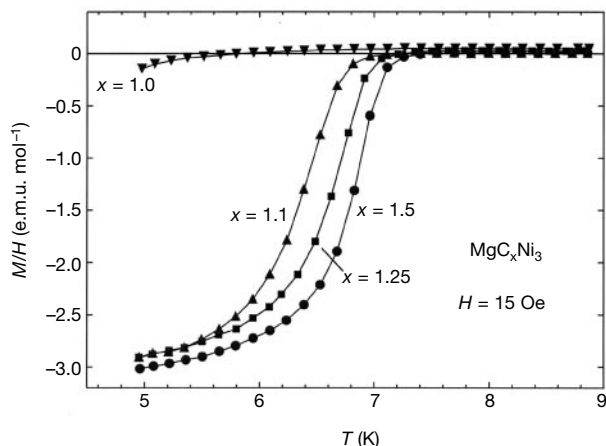


Figure 2 Magnetic characterization of the superconducting transitions for the intermetallic perovskite superconductor of nominal composition MgC_xNi_3 . Samples in the form of loose powders, measured in a Quantum Design PPMS magnetometer under an applied d.c. field of 15 Oe. Data taken on heating after cooling in the absence of a field.

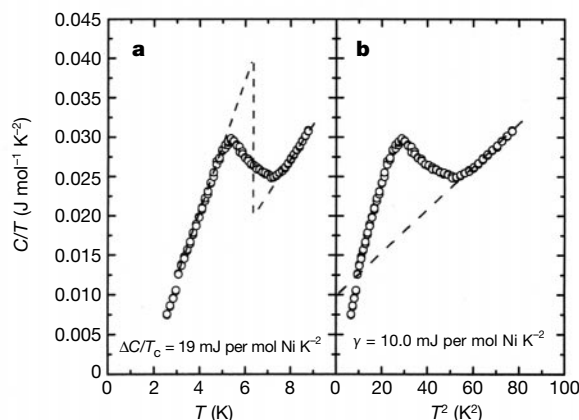


Figure 4 The characterization of the superconducting transition in MgCNi_3 by measurement of the specific heat. Measurements were made using a relaxation method in a commercial calorimeter (Quantum Design). The powder sample was cold-sintered with Ag powder in a $\text{MgCNi}_3:\text{Ag}$ ratio of 3:2 by mass. The contribution of the Ag was measured separately and subtracted. **a**, C/T versus T , to determine the change in the specific heat at the superconducting transition. **b**, C/T versus T^2 , to determine the electronic contribution to the specific heat (γ) by extrapolation to $T = 0$.

like (Ba,K)BiO₃ (see ref. 8, for example) is transparent, the electronic equivalency is not as obvious. An electronic analogy is also possible, however, between these seemingly completely different materials. For the oxide perovskites, the electronic states at the Fermi energy involve holes in the oxygen electronic orbitals. The Ni in the MgCNi₃ perovskite, which is structurally equivalent to O in the oxides, may also be found to have holes in its electronic orbitals at the Fermi energy. Owing to the nearly filled *d* orbitals in elemental nickel, intermetallic compounds of Ni, especially in the presence of electropositive elements such as Mg, often have filled or nearly completely filled Ni *d* states. If this is the case for MgCNi₃, then the conduction will involve holes in Ni *d* states, in an electronic analogy to the holes in the O *p* states in the perovskite oxide superconductors. The fact that superconductivity rather than ferromagnetism occurs in a compound where so much nickel is present is surprising, and suggests that MgCNi₃ is a candidate compound for unconventional superconductivity. □

Received 6 March; accepted 5 April 2001.

1. Rice, T. M. Unconventional superconductors: contrasting cuprates and ruthenates. *Physica C* **341**–**348**, 41–46 (2000).
2. Cava, R. J. *et al.* Superconductivity in the quaternary intermetallic compounds LnNi₂B₂C. *Nature* **367**, 252–253 (1994).
3. Scheil, E. & Huettner, L. Zur Kennetis des Systems Nickel-Kohlenstoff-Magnesium. *Z. Metallk.* **44**, 387–389 (1953).
4. Huettner, L. J. & Stadelmaier, H. H. Ternary carbides of transition metals with aluminum and magnesium. *Acta Metall.* **6**, 367–370 (1958).
5. Nohara, M., Ishiki, M., Takagi, H. & Cava, R. J. Magnetic field dependence of the low-temperature specific heat of the borocarbide superconductor, LuNi₂B₂C. *J. Phys. Soc. Jpn* **66**, 1888–1891 (1997).
6. Kresin, V. Z. & Parchomenko, V. Thermodynamic properties of strong-coupling superconductors. *Sov. Phys. Solid State* **16**, 2180–2184 (1975).
7. Kresin, V. Z. On the critical temperature for any strength of the electron–phonon coupling. *Phys. Lett. A* **122**, 434–438 (1987).
8. Chamberland, B. L. in *Chemistry of Superconducting Materials* 2–104 (ed. Vanderah, T. A.) Noyes, Park Ridge, NJ, 1992).

Acknowledgements

This research was supported by grants from the US National Science Foundation and the US Department of Energy.

Correspondence and requests for materials should be addressed to R.J.C. (e-mail: rcava@princeton.edu).

Atomic-beam alignment of inorganic materials for liquid-crystal displays

P. Chaudhari*, James Lacey*, James Doyle*, Eileen Galligan*, Shui-Chi Alan Lien*, Alesandro Callegari*, Gareth Hougham*, Norton D. Lang*, Paul S. Andry*, Richard John*, Kei-Hsiung Yang*, Minhua Lu*, Chen Cai*, James Speidell*, Sampath Purushothaman*, John Ritsko*, Mahesh Samant†, Joachim Stöhr‡, Yoshiki Nakagawa‡, Yoshimine Katoh‡, Yukito Saitoh‡, Kazumi Sakai‡, Hiroyuki Satoh‡, Shuichi Odahara‡, Hiroki Nakano§, Johji Nakagaki§ & Yasuhiko Shiota§

* IBM Watson Research Centre, PO Box 218, Yorktown Heights, New York 10598, USA

† IBM Almaden Research Centre, 650 Harry Road, San Jose, California 95120-6099, USA

‡ IBM Japan Ltd, Display Business Unit, Yamato-shi, Kenagawa-ken, Japan

§ IBM Japan Ltd, Advanced Manufacturing Engineering, Yasu-gan, Shiga-ken, Japan

The technique used to align liquid crystals—rubbing the surface of a substrate on which a liquid crystal is subsequently deposited^{1–3}—has been perfected by the multibillion-dollar liquid-crystal display industry. However, it is widely recognized that a non-contact alignment technique would be highly desirable for future generations of large, high-resolution liquid-crystal displays. A number of alternative alignment techniques have been

reported^{4–7}, but none of these have so far been implemented in large-scale manufacturing. Here, we report a non-contact alignment process, which uses low-energy ion beams impinging at a glancing angle on amorphous inorganic films, such as diamond-like carbon. Using this approach, we have produced both laptop and desktop displays in pilot-line manufacturing, and found that displays of higher quality and reliability could be made at a lower cost than the rubbing technique. The mechanism of alignment is explained by adopting a random network model of atomic arrangement in the inorganic films. Order is induced by exposure to an ion beam because unfavourably oriented rings of atoms are selectively destroyed. The planes of the remaining rings are predominantly parallel to the direction of the ion beam.

Here we describe a display manufacturing process that breaks away from the method of the past two decades and compare the process with a model for the atomic-scale mechanism of alignment of liquid crystals. Our understanding of the mechanism enabled us to transfer, within a few years, an experimental process to the manufacturing floor. We first discuss the advantages of the new process over rubbing in display manufacturing, and then we compare some experimental data with the model for alignment.

It is widely believed that the current rubbing manufacturing process has serious limitations. It introduces debris, which means that the rubbing machine must be housed in a separate room, in an otherwise clean-room environment, which is more expensive. It is sometimes unreliable: if the rubbing roller degrades unevenly, the effect of a local defect may not be detected until hundreds of flawed displays have been manufactured and subsequently discarded.

Rubbing can also produce an electrostatic discharge that can adversely influence the electronic circuitry just below the surface of the rubbed polyimide thin film. Rubbing can leave streaks, which degrades the quality of the image. This problem has become more acute as the display resolution has increased and the spacer balls, currently used to separate the two glass plates, are replaced by posts. These posts produce shadows as the roller traverses them, resulting in poor alignment of the liquid crystal in the shadowed area that, because of contrast variations, is visible.

There are two reasons why the rubbing process has been very difficult to displace in liquid-crystal manufacturing: much research and development has been focused on refining the existing process



Figure 1 A photograph of a laptop computer showing a 13.3 diagonal extended graphics array (XGA) thin-film transistor liquid-crystal display (DLC) film as the alignment layer. An ion beam was used to produce the alignment.

empirically, and any new technology has to satisfy many technical, reliability, cost, and, increasingly, environmental requirements besides improved alignment of liquid crystals. But we believe the process described here meets these challenges.

We have replaced the polyimide (PI) film, which is printed and then baked in a furnace before rubbing, by a cost-effective, one-step evaporated or sputtered film deposited at room temperature. The vapour-deposition process is not only more cost-effective but also more environmentally friendly than the PI, for which relatively large amounts of organic materials and solvents are needed.

We find that a great variety of optically transparent and insulating films, such as hydrogenated diamond-like carbon (DLC), SiN_x , hydrogenated amorphous silicon, SiC , SiO_2 , glass, Al_2O_3 , CeO_2 , SnO_2 , ZnTiO_2 and InTiO_2 can be used as alignment materials. The films are either amorphous or have a very fine grain size. Of these materials we have examined two films in some detail. These are DLC and SiN_x . We have chosen these two materials because of their wide use in electronics, including the display industry. For example, DLC is widely used in disk storage as the final hard layer on disks, and SiN_x is extensively used in the display industry as an insulator. Selecting materials on this basis has the advantage that tools and a knowledge base exist for large-scale manufacturing. Here we use DLC as the prototypical material.

We have replaced the rubbing roller by a low-energy ion gun, available commercially. There is thus no need to wash a rubbed surface to remove debris or to bake it in a furnace to remove the effects of washing. There are also no reliability issues associated with the unpredictable local degradation of the roller. We have manufactured displays, both the twisted nematic displays (15-inch) used in laptops and the in-plane switching displays (22-inch) used in desktops, that show superior front-of-screen quality to displays produced by the rubbing process, primarily because of the absence of rubbing streaks. The new process is more efficient because it is a nonstop manufacturing process; the rubbing technology needs cloth changes, which cause delays in production. An example of a laptop computer display manufactured with the new process is shown in Fig. 1.

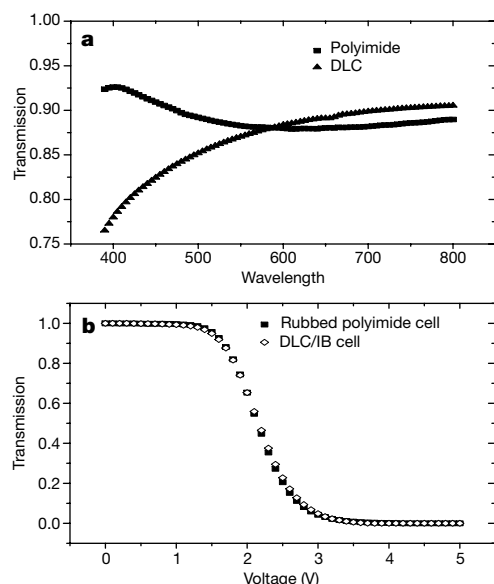


Figure 2 Transmission of light through polyimide and the DLC films on glass substrates and in cells filled with liquid crystals. **a**, Transmission of light through a film of polyimide and DLC deposited on an indium tin oxide coated glass substrates as a function of the wavelength of light. **b**, The transmission of light as a function of applied voltage across twisted nematic liquid-crystal cells made of polyimide and DLC films, as alignment layers.

Carbon films absorb light in the visible range, whereas polyimide (PI) layers are transparent in the visible. We therefore studied DLC films that are 2–10 nm thick. Figure 2a shows transmittance curves as a function of wavelength of light of a DLC and a PI film. Both films are deposited on glass substrates coated with indium tin oxide. The thickness of the DLC film is about 3–4 nm and that of the PI film is about 60 nm. The PI layer seems to have higher transmittance at short wavelengths around 380 nm whereas the carbon films have higher transmittance in the long wavelength range from about 650 to 800 nm.

We have measured the transmission of white light through cells containing liquid crystals (twisted nematics) and either DLC or PI films. The hydrogen content and thickness of the DLC film were adjusted so that the overall transmission of the cell was at least 95% of the PI-containing cells. We obtained typical values of 97%. Also, from data of the type shown in Fig. 2a, we determined the optical constants of our films, and hence calculated the transmission of light through our cells with Berreman's 4 × 4 matrix method⁸. We find the agreement between experiment and calculation to be within 1%.

We show, in Fig. 2b, the transmission of light through two twisted nematic test panels as a function of applied voltage across a cell. The cell spacing is about 5 μm . One of the cells contains a DLC film that has been aligned with an ion beam and the other cell contains a polyimide film aligned by rubbing. The transmission characteristics are very similar for the two panels both at zero and finite voltage. The two sets of data were normalized to unity at zero voltage for the purposes of comparison. Visually, the difference in quality of the two panels is indistinguishable. We also measured the charge retention of these two panels. The ion-beam-aligned DLC had a charge retention of 98%; for the rubbed and aligned test panel it was 97.75% that is, essentially the same.

The ion beam produces alignment with pre-tilt angles ranging from values close to zero to about 10°. A pre-tilt, that is, a tilt of the liquid-crystal axial director away from the plane of the surface, is essential in breaking symmetry and producing a uniform display with high contrast. We found that the pre-tilt angle is a function of

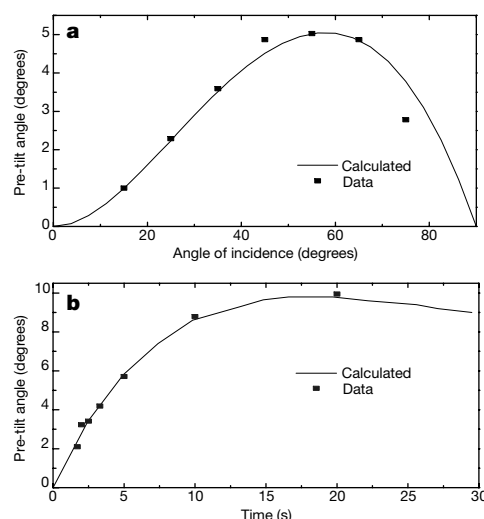


Figure 3 The pre-tilt angle of the liquid crystal as a function of the angle of incidence of the ion beam and time of exposure. All angles are measured from the surface of the substrate. **a**, The pre-tilt angle as a function of the angle of incidence of the ion beam. The film was irradiated with argon ions accelerated to 100 eV, the beam current was 100 mA, and time of exposure was held constant at 30 s. **b**, The pre-tilt angle of the liquid crystal as a function of time of exposure to an argon ion beam. The angle of incidence was held constant at 35 degrees, the acceleration voltage of 200 eV, and the beam current at 100 mA.

the angle of incidence of the ion beam, irradiation times, and ion-beam energy. These findings are similar to our earlier results on the effects of ion beams on polymers⁹. We show in Fig. 3a the pre-tilt angle plotted as a function of the angle of incidence of the ion beam. The data shows a peak in the value of the pre-tilt angle at an incidence angle of the ion beam of about 55°. However, the position of this peak can shift with the time of exposure and the ion beam energy. We show in Fig. 3b the value of the pre-tilt angle as a function of exposure time. This too shows a maximum, whose location changes with the angle of incidence and ion beam energy. It is clear that any desired pre-tilt angle, between approximately 0 and 10° can be obtained by a combination of the angle of incidence of the ion beam and time of exposure.

Another parameter that is important for the functioning of displays is called image sticking. This arises from residual charges that accumulate in a local region as the voltage is left on. When the voltage is removed, the image survives and gradually fades away with time as the charge is dissipated. We have attempted to measure image sticking in a number of twisted nematic cells used in laptops, but so far it has been negligible and less than what we find in rubbed polyimide samples. Image sticking has often been a limiting phenomenon in the acceptance of new ideas to replace the rubbing process.

One interesting phenomenon in rubbing is the ability to overwrite the alignment produced by aligning first in one direction and then rubbing over it in a different direction. This has the advantage of reducing the number of masking steps required in producing a multi-domain display—such a display is characterized by discrete alignment directions of the liquid crystal. We found that the ion beam can also be used to overwrite a previously aligned area. We used this observation to form a two-domain pattern using a single masking step. We first exposed the entire sample to an ion beam and then, using a contact metal mask, selectively exposed the rest of the area in a direction rotated by 180° from the first exposure. We thus fabricated a twisted nematic, two-domain cell in the laboratory. The measured contrast ratio contour, as a function of viewing angle, of such a cell is shown in Fig. 4a. We compared it with the measured contrast ratio contour, as a function of viewing angle, of a single-domain cell shown in Fig. 4b. We note, however, that the contact metal mask is not the best way of producing two domains in a manufacturing environment. A better method would be to use a noncontact metal mask and the properties of ion beams to scan and

produce the second domain. We are working on this possibility in the laboratory. However, for desktop displays, where battery life-time is not an issue, it is possible to use an in-plane switching approach and produce one- or two-domain displays by appropriate design of the switching fields. This is the approach we have taken in producing the high-resolution, two-domain 22-inch display in a manufacturing environment by the process described here. This display currently has the best existing resolution for large displays. It has 3,840 by 2,400 (about 9.2 million) pixels.

We now present a model for the rearrangement of atoms on the surface of an alignment layer produced by a low-energy, collimated ion beam. We assume that the alignment layer is amorphous and isotropic. By isotropic, we mean that the atomic arrangement has no preferred direction before the ion beam is turned on. For the purposes of demonstrating our model, we shall assume that the atomic arrangement can be described by a random network. This is appropriate for covalently bonded solids. However, the precise topology is not important in our modelling. If, for example, the surface is described by a dense random packing of hard spheres, as is generally the case for metallic bonding, we know that there is a transformation that enables us to go from a random network to a dense packing of hard spheres¹⁰; thus the generality of our argument is still valid.

The topology of a network can be described in terms of ring statistics. For example, a random network describing amorphous silicon contains six-fold, five-fold and seven-fold rings. The orientation of these rings, that is, the vector normal to the plane of the rings, is uniform in all directions. The rings present an anisotropic cross-section to the incoming ion beam. For example, if the ion beam is normal to the plane of the ring it has a much larger probability of interaction than if the ion beam is parallel to the plane of the ring. We use this selectivity, as in Darwin's theory of evolution, to destroy rings that present a larger cross-section than others.

After exposure to an ion beam, what is left behind are predominantly those rings whose ring vector is perpendicular to the ion-beam direction. The liquid-crystal molecules are sensitive to the surface arrangement of atoms and hence align along a direction defined by the ion-beam direction projected onto the surface of a sample. The pre-tilt angle of the liquid crystals is generated, in this model, by those rings whose plane is parallel to the direction of the ion beam.

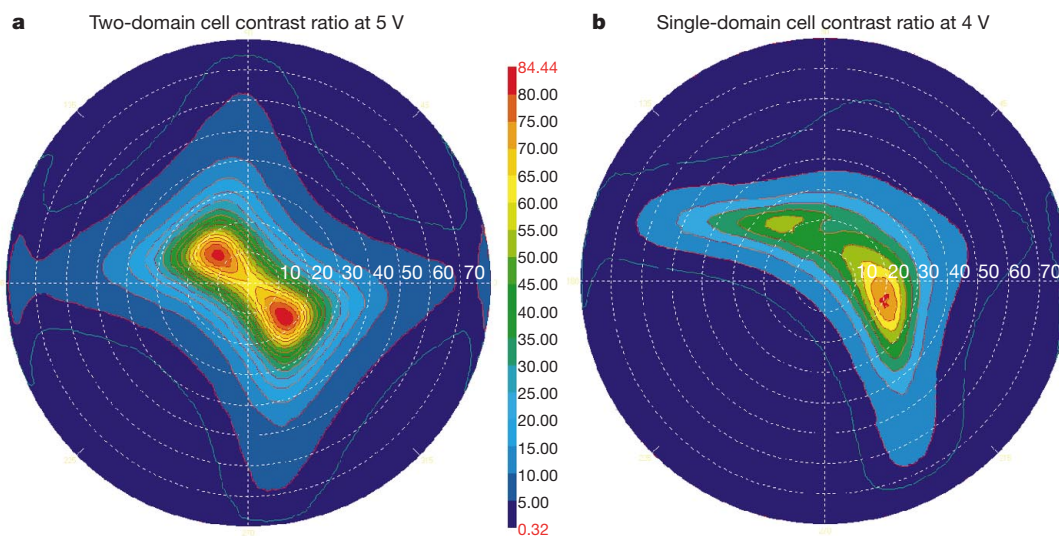


Figure 4 The transmission contrast ratio contours measured as a function of viewing angle for one- and two-domain nematic liquid crystal cells. **a**, Contrast ratio curves obtained from two domains. The second domain was produced using a metal mask with a

feature size of about 180 μm . The metal mask was 75 μm thick and made of molybdenum. It was held in contact with the DLC film during ion beam exposure. **b**, Contrast ratio from a single domain.

A comparison of the calculated and measured values of the pre-tilt angles are shown in Fig. 3a and b. The values of the pre-tilt angles were calculated using the equation described in the Methods section. For purposes of comparison with our experiments, we have adjusted the model calculations to match at one point of the pre-tilt versus ion-beam angle and pre-tilt versus ion-beam exposure time data. We note that the model predicts quite well the functional dependence of the pre-tilt angle with angle and time of exposure. A more detailed description of the model and its comparison with experiments, using both liquid-crystal and X-ray data, will be provided elsewhere (P.C., N.D.L., M.S., J.S. and J. Lüning, unpublished work). □

Methods

Optimization of hydrogen content in DLC films

The DLC films were deposited by plasma-enhanced chemical vapour deposition on glass substrates coated with a conducting indium tin oxide (ITO) film about 300 Å thick. Other substrates, such as Si and quartz, were also used for process comparison. The amorphous (a)-C:H films were deposited using C₂H₂/He and H₂ gas mixtures. Hydrogen was added to the process to increase film transmittance. The hydrogen content of the films was measured by forward recoil scattering. We used a process matrix to establish which process parameters yield higher hydrogen content and thus higher transmittance. In all of these experiments, the substrates were held at room temperature, which is important for cost considerations and also helped to obtain more transparent films than would a high-temperature process.

Ion-beam exposure

Ion beams were produced with a direct-current Kaufman-type ion source using a tungsten filament to supply electrons to the plasma and a plasma-bridge neutralizer to maintain charge neutrality. Argon was introduced into the ion source and plasma-bridge neutralizer using separate gas-flow controllers. The plasma-bridge neutralizer has the advantage of reduced substrate contamination as it is not physically immersed in the ion beam. In contrast, the commonly used tungsten-filament neutralizer, which is placed directly in the ion beam, is subject to sputtering by the energetic argon ions and neutrals impinging on it. Ions with energies in the range of 50–500 eV were extracted for the plasma within the ion source and accelerated toward the substrate. Ion current densities were measured using biased Faraday-cup probes to repel low-energy electrons introduced into the beam by the plasma-bridge neutralizer.

Substrates were mounted on a moving stage that was linearly scanned beneath the ion source. The tray speed was programmable, allowing different ion doses to be applied to the sample when using a fixed ion current density. The incident angle of the impinging ions could be varied by adjusting the angular position of the ion source relative to the substrate.

Calculation of the pre-tilt angle as a function of ion beam conditions

We quantify the model, described qualitatively in the text, by assuming that the destruction probability for a ring is proportional to the beam flux through it, and that we can neglect the effect of the destruction of other rings on that of a given ring. Let us assume a coordinate system with $\mathbf{x}$ normal to the surface and the ion beam incident in the x - y plane towards positive z , with angle of incidence α relative to the z axis. We define a vector $\mathbf{n}$ normal to a given ring, and denote the angle between $\mathbf{n}$ and the z axis as θ , and the azimuthal angle in the x - y plane as ϕ .

We denote by $\langle \theta \rangle$ the mean value of θ over the distribution of rings remaining on the surface at a time t , with beam exposure starting at $t = 0$. At $t = 0$, $\langle \theta \rangle = \pi/2$, because the beam distribution starts out isotropic. We take the product, denoted by W , of $\pi/2 - \langle \theta \rangle$ and the fraction of rings remaining on the surface at time t to be a measure of the pre-tilt angle; given the symmetry of the problem, we find that

$$W = \frac{1}{\pi} \int_0^\pi \left(\frac{\pi}{2} - \theta \right) \sin \theta \, d\theta \int_0^{2\pi} \exp(-pt \sin \alpha |\sin \alpha \sin \theta \cos \phi - \cos \alpha \cos \theta|) d\phi$$

where p is proportional to the beam intensity, ring area, and destruction probability per unit time per unit beam flux through a ring. Using this equation, we have calculated the variation in the pre-tilt angle with the ion-beam dose and its angle of incidence. For details of the model, see P.C., N.D.L., M.S., J.S. and J. Lüning, unpublished work).

Received 30 November 2000; accepted 19 March 2001.

- DeGennes, P. G. & Prost, J. *The Physics of Liquid Crystals* 161 (Clarendon, Oxford, 1993).
- Stöhr, J. *et al.* Microscopic origin of liquid crystal alignment on buffed polymer surfaces. *Macromolecules* **31**, 1942–1946 (1998).
- Stöhr, J. & Samant, M. G. Liquid crystal alignment by rubbed polymer surfaces: A microscopic bond orientation model. *J. Electron Spectrosc. Rel. Phenom.* **98–99**, 189–207 (1999).
- Aoyama, H., Yamazaki, Y., Matsura, M., Mada, H. & Kobayashi, S. Alignment of liquid crystals on the stretched polymer films. *Mol. Cryst. Liq. Cryst.* **72**, 127–132 (1981).
- Ikeno, Y. *et al.* Electrooptic bistability of a ferroelectric liquid crystal device using polyimide Langmuir–Blodgett orientation films. *Jpn J. Appl. Phys.* **27**, L475–L476 (1998).

- Janning, J. L. Thin film surface orientation for liquid crystals. *Appl. Phys. Lett.* **21**, 173–174 (1972).
- Schadt, M., Seiberle, H. & Schuster, A. Optical patterning of multi-domain liquid-crystal displays with wide viewing angles. *Nature* **381**, 212–215 (1996).
- Berremann, D. W. Optics in stratified and anisotropic media: 4 × 4-matrix formulation. *J. Opt. Soc. Am.* **62**, 502–510 (1972).
- Chaudhari, P., Lacey, J., Lien, S. A. & Speidell, J. L. Atomic beam alignment of liquid crystals. *Jpn J. Appl. Phys.* **37**, L55–L56 (1998).
- Chaudhari, P., Graczyk, J. F., Henderson, D. & Steinhardt, P. Transformation between random network and dense random-packed models for amorphous solids. *Phil. Mag.* **31**, 727–732 (1975).

Acknowledgements

We thank T. Hashimoto for his encouragement and T. Ueki, H. Yamaguchi and M. Mitsuhashi for their sustained support. P.C. thanks N. Okazaki for describing this technology to colleagues in Japan. G. Thompson photographed Figure 1.

Correspondence and requests for materials should be addressed to P.C. (e-mail: chaudha@us.ibm.com).

Size-dependent control of the binding of biotinylated proteins to streptavidin using a polymer shield

Zhongli Ding, Robin B. Fong, Cynthia J. Long, Patrick S. Stayton & Allan S. Hoffman

Department of Bioengineering, University of Washington, Seattle, Washington 98195, USA

Many medical and biotechnological processes rely on controlling and manipulating the molecular-recognition capabilities of proteins^{1–4}. This can be achieved using small molecules capable of competing for protein binding or by changing environmental parameters that affect protein structure and hence binding. An alternative is provided by stimuli-responsive polymers that change reversibly from a water-soluble expanded coil to a water-insoluble collapsed globule upon small changes in temperature, pH or light intensity: when attached to proteins in the vicinity of their binding sites, they reversibly block and release small ligands^{1,5–7}. Here we show how this approach can be extended to achieve size-selective binding of large, macromolecular ligands. We use the thermally responsive polymer poly(*N,N*-diethylacrylamide) (PDEAAm), and attach it to the protein streptavidin approximately 20 Å from the binding site for biotinylated proteins. Below the lower critical solution temperature of PDEAAm, the polymer is in its extended state and acts as a 'shield' to block the binding of large biotinylated proteins; above this temperature, it collapses and exposes the binding site, thereby allowing binding. We find that the degree of shielding depends on both the size of the biotinylated protein and the size of PDEAAm, suggesting that 'smart' polymer shields could be tailored to achieve a wide range of size-dependent ligand discrimination for use in affinity separations, biosensors and diagnostics technologies.

We have previously described the conjugation of stimuli-responsive polymers to genetically engineered sites near the binding site (or pocket) of streptavidin, yielding 'affinity switches' that control the binding and release of the small ligand biotin through small changes in environmental conditions^{1,5–7}. The stimuli-responsive polymers reversibly cycle between an expanded coil and a collapsed globule in response to small changes in temperature, pH, or light, with a change in hydrodynamic radius of about three times⁸. In our previous work, the polymer was attached very close to one of

streptavidin's four biotin-binding pockets, and controlled the binding of biotin binding at this pocket.

Here we exploit the approximate C_{222} symmetry of the streptavidin tetramer, which gives rise to two distinct biotin-binding faces, with each binding face presenting two biotin-binding pockets separated by ~ 20 Å. This allows us to investigate the shielding activities of stimuli-responsive polymers attached at positions more distant from the biotin-binding site, as a function of the size of the ligand binding to streptavidin. Different molecular masses of the thermally responsive polymer, PDEAAm, were conjugated to a E51K/N118K streptavidin mutant and immobilized on magnetic microbeads (see Methods). This mutant was designed to enhance conjugation efficiency by localizing reactive primary amine groups near the biotin-binding pocket. Radiolabelled biotin-binding experiments showed that only one polymer could be conjugated per streptavidin-binding face.

The conjugate made from 12.8-kDa PDEAAm and E51K/N118K streptavidin (E51K/N118K–PDEAAm-12.8k) shielded the 20-Å-distant biotin-binding site against biotinylated proteins. Figure 1 shows that the amount of biotinylated bovine serum albumin (B-BSA) bound to this conjugate on the microbeads is temperature dependent, with low binding below 20 °C, followed by a sharp increase upon increase of temperature. Above 30 °C, the quantity of B-BSA bound to the conjugate is the same as that bound to the unconjugated E51K/N118K streptavidin. In the control, the binding of B-BSA to the unconjugated E51K/N118K streptavidin is high at all temperatures. The temperature dependence of the shielding effect thus closely corresponds to the phase transition behaviour of the polymer, which displays a lower critical solution temperature (LCST) of about 24 °C in the buffer solution.

The shielding activity of the streptavidin–PDEAAm-12.8k conjugate depends strongly on the size of the biotinylated protein target. A small (6.2 kDa) biotinylated protein, B-protein G, binds efficiently both above and below the LCST of PDEAAm (Fig. 1). B-protein G is not sterically shielded by the hydrated PDEAAm, and

can access the nearby biotin-binding site whether the PDEAAm is extended or collapsed (Fig. 2). We also investigated a much larger target protein, biotinylated immunoglobulin- γ (B-IgG; 150 kDa) and found its binding to the conjugate was low over the entire temperature range. Due to the large size of B-IgG, its access to the binding site is effectively shielded even by the collapsed 12.8-kDa PDEAAm (Fig. 2).

Because of steric restrictions, each binding face of the unconjugated E51K/N118K streptavidin mutant can only bind one B-BSA. Thus, the low binding efficiency of B-BSA to E51K/N118K–PDEAAm-12.8k below the LCST of the conjugate (Fig. 1) arises because the extended, hydrated polymer coil obstructs access of B-BSA to the nearby binding pocket of streptavidin (Fig. 2). When the polymer is collapsed above the LCST, the binding site where the polymer is directly conjugated still remains shielded, while access to the nearby binding site at 20 Å distance is no longer obstructed, allowing binding of the B-BSA.

To confirm this shielding mechanism, we derivatized PDEAAm oligomers with biotin at one end, and showed that the biotinylated oligomer was able to associate with an average of only one binding site per binding face of the unconjugated E51K/N118K streptavidin mutant in solution. And when complexing one biotinylated 12.8-kDa PDEAAm per immobilized E51K/N118K streptavidin tetramer, the blocking of B-BSA as a function of temperature was similar to that seen with the immobilized conjugate (Fig. 3). This control experiment confirms that conjugation, as well as complexation, of a single smart polymer chain at a distance of about 20 Å to the nearby biotin-binding site can be used to reversibly shield and unshield that site.

The dependence of the shielding activity on the size of the biotinylated target proteins suggested that the size-selectivity of the polymer shields could be 'fine-tuned' by using polymers of different sizes. To test this, we prepared conjugates with PDEAAm oligomers of different molecular masses. At 10 °C, which is below the LCST of the PDEAAm, the amount of B-BSA bound to the

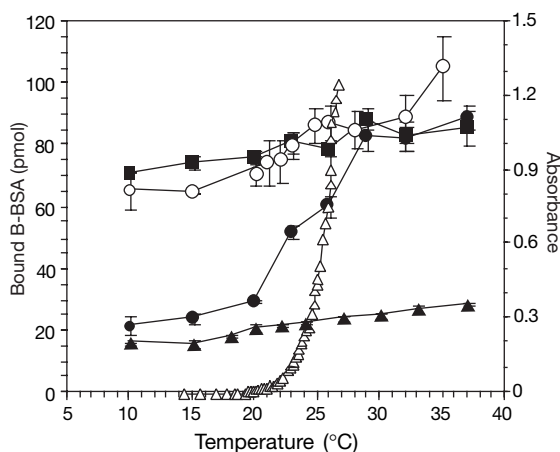


Figure 1 Shielding of biotinylated protein binding to E51K/N118K–PDEAAm conjugates. The binding of biotinylated BSA (B-BSA) to the E51K/N118K–PDEAAm-12.8k conjugate (filled circles) rises significantly as the temperature rises through the LCST (~ 23 °C), whereas the binding of B-BSA to unconjugated E51K/N118K streptavidin (filled squares) is efficient at all temperatures studied. The much smaller B-protein G binds to the E51K/N118K–PDEAAm-12.8k conjugate (open circles) efficiently over the whole temperature range investigated, while the binding of the large protein B-IgG to the E51K/N118K–PDEAAm-12.8k conjugate (filled triangles) is shielded over the entire temperature range. Under the binding conditions used (see Methods), the E51K/N118K–PDEAAm-12.8k conjugate has an LCST of ~ 23 °C (open triangles).

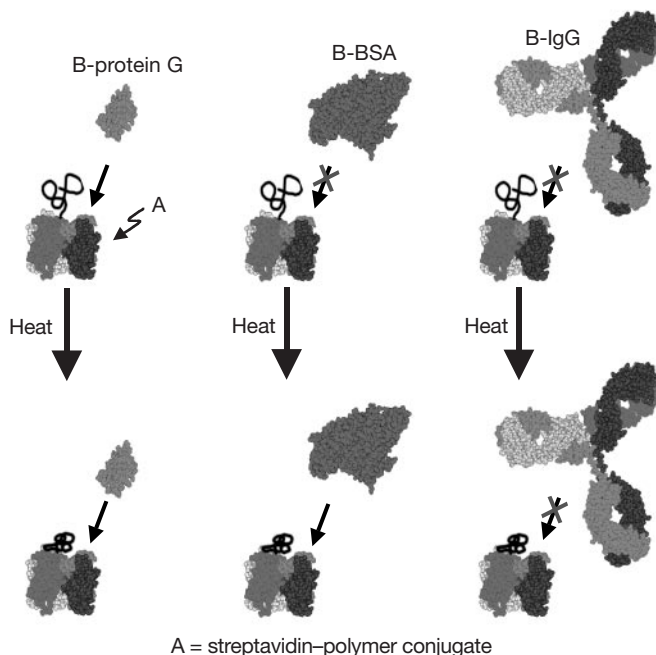


Figure 2 Schematic summary of the temperature sensitivity of the binding of biotinylated protein to streptavidin conjugates. The protein models were generated from Protein Data Bank files, with human albumin serving as a close analogue. The proteins are thus represented in proportion to their molecular sizes.

streptavidin–PDEAAm conjugates increases as the molecular mass of the conjugated polymer decreases (Figs 3, 4). This shows that the shielding activity can indeed be tuned by controlling the molecular mass of the conjugated polymer and supports the steric hindrance model, where smaller polymer chains provide less steric hindrance and allow greater access of biotinylated proteins to the adjacent binding site. As a further test of this model and of the size-selective shielding capabilities, we measured the ability of the streptavidin–PDEAAm–12.8-kDa conjugates to discriminate between B-protein G and B-BSA together in solution. The PDEAAm shield blocked B-BSA association, while allowing the B-protein G to bind to the same level as seen when it was the only protein in solution (data not shown). Thus, the conjugated smart polymer shields can also be used to discriminate amongst mixtures of proteins on the basis of size.

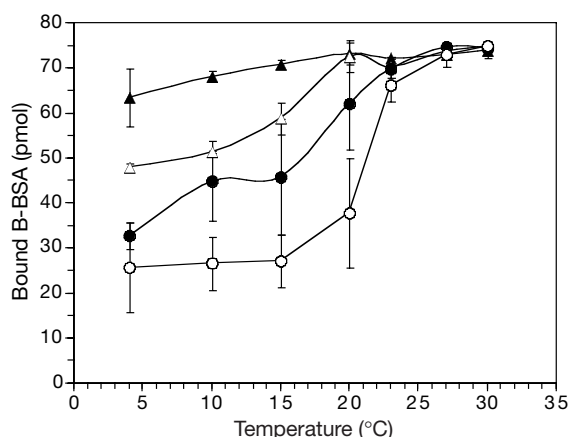


Figure 3 Shielding activity as a function of temperature and molecular mass of the PDEAAm polymer. Shown are the binding activities of B-BSA to immobilized E51K/N118K streptavidin that is either unconjugated, or forming an affinity-bound complex with biotinylated PDEAAm (B-PDEAAm) of different molecular masses. The graph shows the binding of B-BSA to E51K/N118K streptavidin (filled triangles) and complexes of E51K/N118K streptavidin and B-PDEAAm with molecular masses of 2.0 kDa (open triangles), 6.7 kDa (filled circles) and 12.8 kDa (open circles).

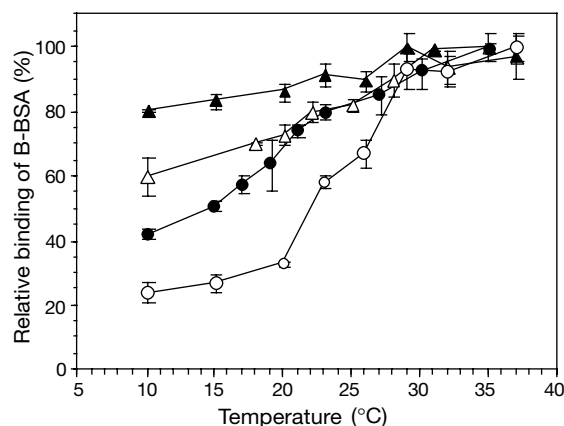


Figure 4 Effect of molecular mass of PDEAAm on the shielding efficiency. The graphs show the binding of B-BSA to E51K/N118K streptavidin (filled triangles) and covalent conjugates of E51K/N118K–PDEAAm with molecular masses of 2.0 kDa (open triangles), 6.7 kDa (filled circles) and 12.8 kDa (open circles). To compare the effect of the different molecular-mass polymers on the shielding efficiency, we show relative binding of B-BSA, which is obtained by normalizing the amount of bound B-BSA to the maximum binding for each sample.

The small ligand biotin can access the binding pocket where the polymer is directly conjugated when the polymer is in the expanded state, allowing the collapse or expansion of the polymer to modulate association and/or dissociation^{5–7}. However, the binding of the macromolecular biotinylated proteins is always blocked at the binding pocket where the polymer is conjugated. The biotinylated proteins thus provide a model system for developing the concept of the molecular shield, where the polymer is conjugated at a nearby but not directly adjacent site (in this case, approximately 20 Å away). The expanded polymer then becomes the active blocking state, because it can be engineered to be an appropriate size for sterically blocking access to the distant active site, whereas collapse unmasks the blocking activity in a size-dependent manner. This comparison is shown in Fig. 5.

The general approach should in principle be applicable to any protein active site, provided the polymer is conjugated at an appropriate distance from the active site so that its collapse unmasks that site. It should be possible to engineer the specificity and selectivity of these molecular shields, by matching the sizes of the polymer to those of the targets that undergo affinity binding. We expect that this approach to achieving size-selective shielding of protein active sites might find use in biotechnology applications. For example, control of the polymer coil size within or near the active site of a polymer–enzyme conjugate through temperature, pH or light signals, could favour the enzymatic transformation of smaller substrates, due to ‘gating’ (shielding) of large substrates. It might also be possible to use this approach to control access of pro-drugs to enzyme active sites, so that the active sites are obstructed while therapeutic enzymes travel to their target site, and opened on arrival. Similarly, it might be possible to efficiently separate mixtures of peptide and protein ligands on the basis of their binding affinity and size. Last, the reversible blocking and opening of immobilized protein arrays to target molecules such as peptides, DNA or photoactive molecules could open avenues to nanopatterning and information read/write technologies. □

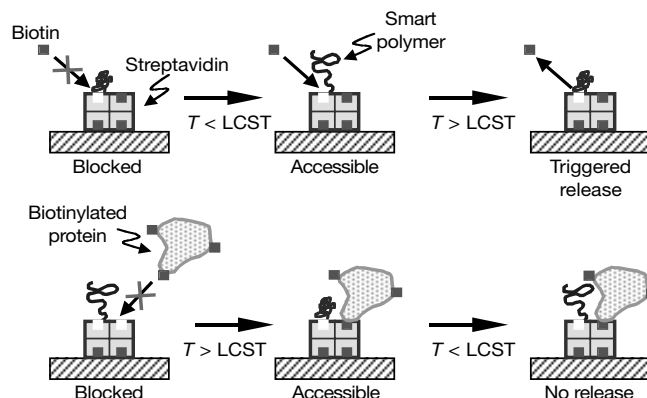


Figure 5 Schematic model of different shielding mechanisms. The model illustrates the shielding effects of a conjugated smart polymer on the binding of a small ligand (biotin) and a large macromolecular ligand (a biotinylated protein) to streptavidin. The polymer is in both cases conjugated to one binding site on the exposed face of an immobilized streptavidin molecule. (Only one polymer chain may be conjugated per binding face, due to steric hindrance). In the case of biotin association, the collapsed polymer blocks biotin access to the same site where the polymer is conjugated, but permits association at that site when the polymer is rehydrated and extended away from the site. In the case of the larger biotinylated protein, the conjugated polymer always blocks binding to the site where the polymer is conjugated, whether the polymer is collapsed or extended. The nearby binding site is exposed and accessible for binding of a biotinylated protein only when the polymer is collapsed, and provided the protein is small enough. Reversal of the stimulus can lead to release (depletion) of the biotin from the binding site where the polymer is conjugated^{6,7} but there is no reversal expected for the biotinylated protein because the polymer is conjugated to the subunit 20 Å away from the binding site.

Methods

PDEAAm synthesis

PDEAAm was prepared by group transfer polymerization⁹ to ensure a narrow molecular-mass distribution. Tetrabutylammonium acetate was used as the catalyst and 1-methoxyl-1-(trimethylsiloxy)-2-methyl-1-propene as the initiator. In the final stage of polymerization, a hydroxyl group was incorporated into one end of each polymer chain by adding a capping agent, 2-(trimethylsiloxy)ethyl methacrylate. The trimethylsiloxy group was subsequently removed by hydrolysis, to generate the –OH end group. The hydroxyl end group was subsequently derivatized to a vinyl sulphone group by reacting with divinyl sulphone^{6,7}. The molecular mass and polydispersity of the polymer were determined in tetrahydrofuran, using gel permeation chromatography calibrated with polystyrene molecular-mass standards. All of the polymers have a molecular-mass polydispersity of less than 1.2.

Conjugation

Conjugation of PDEAAm to E51K/N118K streptavidin was performed at pH 9.5, 4 °C for 16 h. Three thermally induced precipitations of the conjugate were conducted to remove any unconjugated E51K/N118K streptavidin, which remained in the supernatant. Iminobiotin affinity chromatography was employed to separate the conjugate from the unreacted free polymer¹⁰. The purified streptavidin–PDEAAm conjugate was then immobilized on magnetic microbeads for the biotinylated-protein binding assays.

Binding assays

The conjugates and, as a control, the unconjugated E51K/N118K streptavidin were immobilized on magnetic microbeads and suspended in 100 mM sodium phosphate buffer, pH 8, containing 0.2 wt% of BSA. The suspensions were incubated in a 10 °C water bath for 1 h before addition of the biotinylated protein. They were mixed and then further incubated at 10 °C for 30 min to reach binding equilibrium. The magnetic beads were separated and the fluorescence intensities (excitation wavelength, 494 nm; emission wavelength, 520 nm) of the supernatant were measured versus control solutions without magnetic microbeads. The solutions were then incubated at a higher temperature and the same operations were repeated. One nmol of either E51K/N118K–PDEAAm conjugate or E51K/N118K was used for each assay.

Determination of LCST

The LCST of the E51K/N118K–PDEAAm-12.8k conjugate was determined in 100 mM sodium phosphate buffer, pH 8.0, containing 0.2 wt% of BSA, by measuring absorbance at 500 nm versus temperature. The LCST is defined as the temperature where the light absorbance is 10% of the maximum value.

Synthesis of biotinylated PDEAAm

A primary amino group at the end of the PDEAAm chain was reacted with sulph-NHS-LC-biotin (from Pierce). B-PDEAAm was then complexed to E51K/N118K streptavidin, which was immobilized on magnetic beads via the interaction of biotin and streptavidin.

Received 29 December 2000; accepted 12 February 2001.

1. Fong, R. B., Ding, Z., Long, C. J., Hoffman, A. S. & Stayton, P. S. Thermoprecipitation of streptavidin via oligonucleotide-mediated self-assembly with poly(N-isopropylacrylamide). *Bioconj. Chem.* **10**, 720–725 (1999).
2. Wilchek, M. & Bayer, E. A. *Avidin–Biotin Technology* (Academic, New York, 1990).
3. Schlosser, M., Hahmann, J., Ziegler, B., Augstein, P. & Ziegler, M. Sensitive monoclonal antibody-based sandwich ELISA for determination of the diabetes-associated autoantigen glutamic acid decarboxylase GAD65. *J. Immunoass.* **18**, 289–307 (1997).
4. Bloch, B. Biotinylated probes for in situ hybridization histochemistry: Use for mRNA detection. *J. Histochem. Cytochem.* **41**, 1751–1754 (1993).
5. Stayton, P. S. et al. Control of protein–ligand recognition using a stimuli-responsive polymer. *Nature* **378**, 472–474 (1995).
6. Bulmus, V., Ding, Z., Long, C. J., Stayton, P. S. & Hoffman, A. S. Design, synthesis and site-specific conjugation of a pH- and temperature-sensitive polymer to streptavidin for pH-controlled binding and triggered release of biotin. *Bioconj. Chem.* **11**, 78–83 (1999).
7. Ding, Z. et al. Temperature control of biotin binding and release with a streptavidin-poly(N-isopropylacrylamide) site-specific conjugate. *Bioconj. Chem.* **10**, 395–400 (1999).
8. Wu, C. & Wang, X. H. Globule-to-coil transition of a single homopolymer chain in solution. *Phys. Rev. Lett.* **80**, 4092–4094 (1998).
9. Sogah, D. Y., Hertler, W. R., Webster, O. W. & Cohen, G. M. Group transfer polymerization. Polymerization of acrylic monomers. *Macromolecules* **20**, 1473–1488 (1987).
10. Hofmann, K., Wood, S. A., Brinton, C. C., Montibeller, J. A. & Finn, F. M. Iminobiotin affinity columns and their application to retrieval of streptavidin. *Proc. Natl Acad. Sci. USA* **77**, 4666–4668 (1980).

Acknowledgements

We thank R. Clark for providing protein G, and N. Murthy for help with PDEAAm synthesis. This work was supported by the National Institutes of Health.

Correspondence and requests for materials should be addressed to A.S.H. (hoffman@u.washington.edu) or P.S.S. (stayton@u.washington.edu).

Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times

An Zhisheng*, John E. Kutzbach†, Warren L. Prell‡ & Stephen C. Porter§

* State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Box 17, Xi'an 710054, China

† Center for Climatic Research, Institute for Environmental Studies, University of Wisconsin–Madison, 1225 W. Dayton Street, Madison, Wisconsin 53706, USA

‡ Geological Sciences, Box 1846, Brown University, Providence, Rhode Island 02912-1846, USA

§ Quaternary Research Center, Box 351310, University of Washington, Seattle, Washington 98195, USA

The climates of Asia are affected significantly by the extent and height of the Himalayan mountains and the Tibetan plateau^{1–4}. Uplift of this region began about 50 Myr ago, and further significant increases in altitude of the Tibetan plateau are thought to have occurred about 10–8 Myr ago^{4,5}, or more recently. However, the climatic consequences of this uplift remain unclear. Here we use records of aeolian sediments from China^{6,7} and marine sediments from the Indian^{8–10} and North Pacific oceans¹¹ to identify three stages of evolution of Asian climates: first, enhanced aridity in the Asian interior and onset of the Indian and east Asian monsoons, about 9–8 Myr ago; next, continued intensification of the east Asian summer and winter monsoons, together with increased dust transport to the North Pacific Ocean¹¹, about 3.6–2.6 Myr ago; and last, increased variability and possible weakening of the Indian and east Asian summer monsoons and continued strengthening of the east Asian winter monsoon since about 2.6 Myr ago. The results of a numerical climate-model experiment, using idealized stepwise increases of mountain–plateau elevation, support the argument that the stages in evolution of Asian monsoons are linked to phases of Himalaya–Tibetan plateau uplift and to Northern Hemisphere glaciation.

Continuous sedimentary records of Asian climate are found in China and in marine cores from the Indian and North Pacific oceans (Fig. 1). The planktonic foraminifer *Globigerina bulloides* and upwelling radiolaria from ODP site 722 (Fig. 2) are indices of coastal upwelling in the Arabian Sea and thus of southwesterly wind strength during the Indian summer monsoon^{8,10}. Although carbonate dissolution (often associated with high productivity) reduces the magnitude of the *G. bulloides* index at certain times (Fig. 2), the composite radiolarian and *G. bulloides* records show strengthening of upwelling about 9–8 Myr ago and relatively continuous upwelling thereafter. Magnetic susceptibility flux from ODP site 758 (ref. 9), representing sea-level-mediated fluvial transport from the Ganges and other river systems draining the southern side of the Himalayan–Tibetan orogen, increases about 9 Myr ago. Significantly, new basal dates from the aeolian ‘Red Clay’ sediments on the Chinese Loess plateau (Figs 1, 2) indicate onset of aeolian dust accumulation at about 7.6 Myr ago at Zhaojiachuan (35° 53' N, 107° 58' E), 8.0 Myr ago at Chaona (35° 06' N, 107° 12' E), and as early as 8.3 Myr ago at Jiaxian (38° 16' N, 110° 5' E) (Fang, X.M. and Qiang, X.K., personal communication). Records from North Pacific ODP sites 885 and 886, which accumulated wind-blown dust from Asia, show a major dust peak about 8–7 Myr ago¹¹. The change in oxygen isotope composition of soil carbonates in Pakistan about 9–8 Myr ago¹² (Fig. 2), inferred changes in vegetation from C₃ (forests) to C₄ (grasses) in Pakistan beginning about 8 Myr ago¹³, and a change from mixed needle-leaf and broad-leaf forests to grassland vegetation along the northeastern margin of the Tibetan plateau about 8.5 Myr ago¹⁴, all imply increasing seasonality by about 8 Myr

ago, with most precipitation in summer. These widely distributed observations can be interpreted as signalling an environmental response to a major phase of Himalaya–Tibetan plateau uplift about 9–8 Myr ago. This response is broadly consistent with the climate changes produced in our climate-model experiment. This

experiment used highly idealized stages of elevation history for the Himalayan–Tibetan region; that is, going from relatively small areas with elevations above 1,000 m and maximum elevations between 1,700 and 2,700 m (stages HT-1 and HT-2), with a weak Asian summer monsoon circulation and relatively low summer monsoon

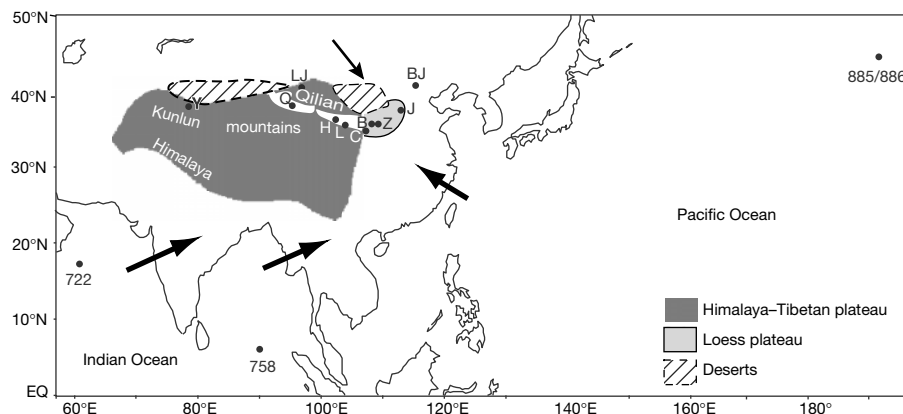


Figure 1 Locations of geographic features and terrestrial and marine records. B.J., Beijing; B, Z and J, locations of the Bajiazui, Zhaojiachuan and Jiaxian sections from the Loess plateau, respectively; C, Chaona site; Y, Yecheng site; L.J., Lao Junmiao section; Q and L,

the Qaidam and Linxia basins; and H, Haiyuan. The bold arrows (narrow arrow) indicate generalized wind directions of the summer (winter) monsoon.

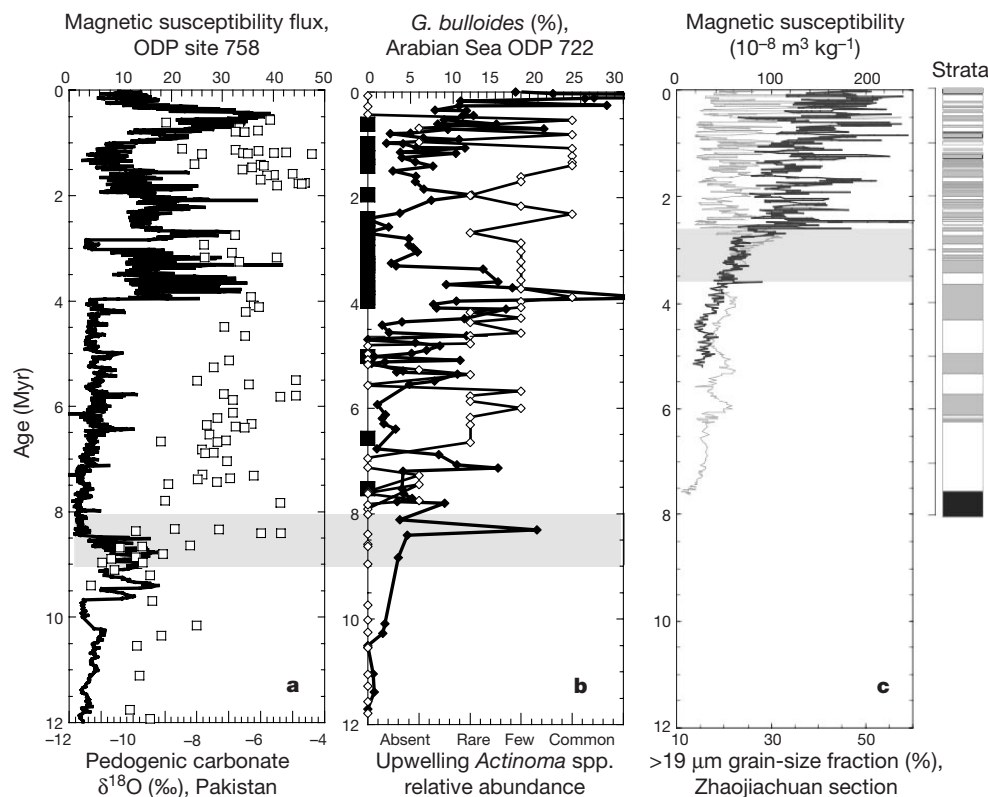


Figure 2 Terrestrial and marine records from the Chinese Loess plateau and the southern margin of Asia. The shaded zones at ~9–8 and 3.6–2.6 Myr ago indicate times of change detailed in the text. **a**, The magnetic susceptibility flux (SI units times sedimentation rate, solid line) from ODP Site 758 (Fig. 1) reflects the sea-level-mediated terrigenous flux to the Bay of Bengal⁹. The $\delta^{18}\text{O}$ of soil carbonates (open squares) in Pakistan¹² reflects increased aridity or a change in the precipitation source about 9–8 Myr ago. The terrestrial record of monsoon influence remains high but variable while the fluvial flux to the Bay of Bengal is low between 8 and 4 Myr ago. This difference is attributed to higher sea level and less transport to the deep sea during the late Miocene. **b**, The abundance of upwelling fauna at ODP Site 722 in the Arabian Sea (Fig. 1). Revised percentage data for planktonic foraminifer *G. bulloides* (filled diamonds) and relative

abundance of radiolarian *Actinoma* spp.⁸ (open diamonds). Intervals of poor carbonate preservation are indicated by filled squares on the time axis. The low *G. bulloides* values from 8–0 Myr ago are mostly attributed to poor carbonate preservation⁸. **c**, The stratigraphy and time series of magnetic susceptibility⁷ (thin line) and >19 μm grain-size fraction (thick line) from the Zhaojiachuan section on the Loess plateau. The chronology is based on the polarity boundary ages^{7,16}, and was obtained by interpolation from a sedimentation model¹⁵ using the >19 μm grain-size fraction and a magnetic susceptibility model⁷ in the basal part where grain-size data are not available. The white, grey, and dark shaded patterns in the simplified stratigraphy column represent loess or loess-like sediment, palaeosols and bedrock, respectively.

precipitation, to a much larger area with elevations above 1,000 m and maximum elevations of 5,700 m (stage HT-3, a stage we associate with the Late Miocene time), with a strong Asian monsoon circulation and increased summer monsoon precipitation (Fig. 3). These changes in continent-scale monsoon circulation are caused primarily by large increases in sensible heating and latent heating (precipitation) that are focused over or along the slopes of the high plateau^{1–4}. In central Asia, precipitation decreases^{2,3} (Fig. 3).

The onset of aeolian deposition in China about 8 Myr ago resulted in long, continuous terrestrial records at Zhaojiachuan and Bajiazui (35° 53' N, 107° 27' E)^{6,7}, within the largest platforms on the Loess plateau (Fig. 1), and in an area that is very sensitive to variations in the east Asian summer and winter monsoons¹⁵. These sequences (Fig. 4) consist of two parts: the upper part corresponds to the well-known loess–palaeosol sequence of Luochuan, aged ≤ 2.6 Myr (ref. 16), which has been correlated with deep-sea sedimentary records; the lower Red Clay sequence consists of inter-layered light-red to reddish-yellow silty loess and light-red to brownish-red palaeosols, and mantles a surface with variable relief and different ages. The quasi-normal grain-size distribution, the 40–60% silt fraction, and other chemical and physical characteristics, indicate an aeolian origin for the sediment^{6,7,17}. Overall southeastward fining of the loessic silt is consistent with north-westerly winter winds. The degree of pedogenesis of the palaeosols, reflected in their colour, texture, and abundance of pedogenic

calcareous nodules, increases southeastward towards regions of increasing summer-monsoon precipitation.

Several indicators of summer and winter monsoon strength have been developed from these loess–palaeosol sequences. The high positive correlation between the frequency-dependent magnetic susceptibility, which can be used to identify ferromagnetic grain size, and the magnetic susceptibility of Red Clay samples indicates that the susceptibility depends mainly on ultra-fine-grained ferromagnetic minerals formed *in situ* during pedogenesis¹⁸. Therefore, susceptibility records in the lower sequence—which have ferromagnetic minerals and magnetic properties similar to those of the overlying loess and palaeosols, and similar susceptibility records to those in the upper sequence—are indices of summer monsoon precipitation¹⁷ (Fig. 4). The overall strong correlation of the magnetic susceptibility series with an independently derived Rb/Sr time-series emphasizes that both indices are measures of summer monsoon strength (Fig. 4). During weathering, Rb is relatively stable, whereas Sr is relatively mobile; therefore, an increased Rb/Sr ratio indicates increased weathering and pedogenesis, and a strong monsoon¹⁹. The coarse-grain fraction and the Al flux (Fig. 4) are indices of winter monsoon strength¹⁵ and the degree of aridity in dust source regions¹¹, respectively.

Based upon the temporal changes of these monsoon indices (Fig. 4), we subdivide the period 6–2 Myr ago into three intervals. The period from 6 to about 3.6 Myr ago shows considerable

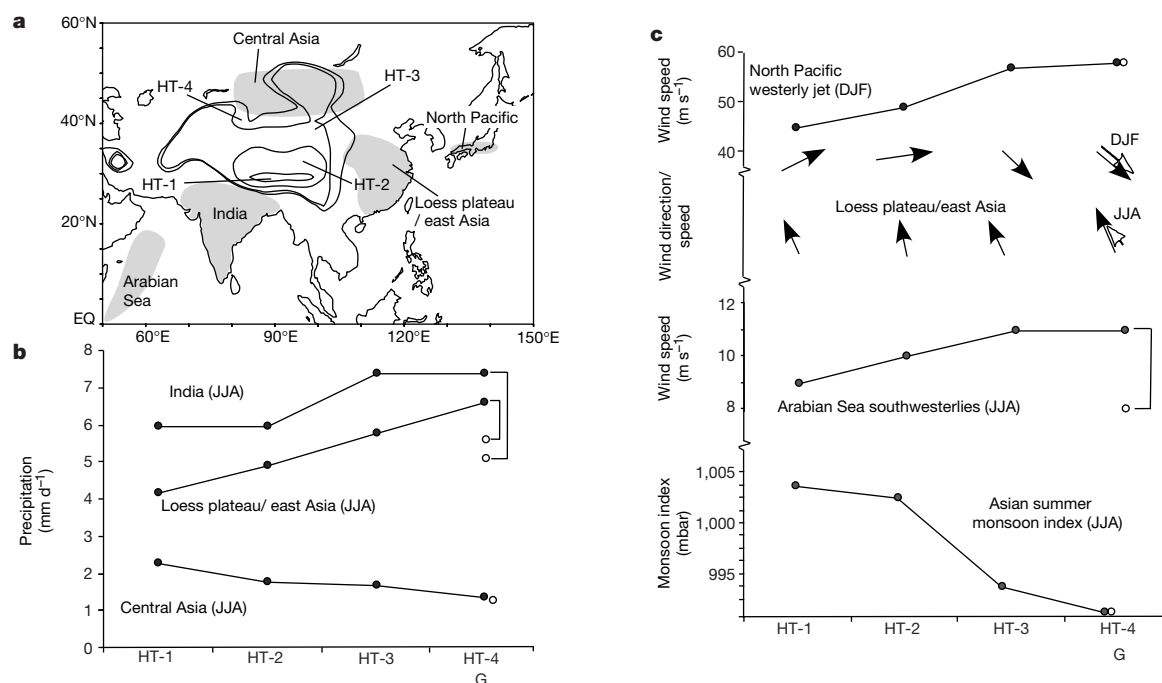


Figure 3 Climate indices from an experiment with four idealized stages of Himalaya–Tibetan plateau elevation (HT-1 to HT-4) and one glacial maximum stage (G) made with the NCAR climate model CCM3. **a**, Areas (in grey) for which climate indices are summarized (below), and approximate boundaries of the idealized topography stages with elevations higher than 1,000 m outlined: HT-1, small elevated region, with maximum elevation less than 1,700 m; HT-2, Himalaya and Tibetan plateau of limited north–south and east–west extent with maximum elevation 2,700 m; HT-3, Himalaya and Tibetan plateau considerably extended to the north and west with maximum elevation 5,700 m; and HT-4, modern, with extension of the plateau along the eastern and northern margins and maximum elevation 5,700 m. The elevations used in the climate model reflect a smoothing of the topography consistent with the spatial resolution of the climate model, and are significantly lower than the observed or estimated elevations. **b**, The June–July–August (JJA) precipitation for India, the Loess plateau/east Asia, and central Asia, for four simulations (HT-1 to HT-4) with progressive increase in mountain–plateau elevation and one simulation (G) with glacial-age modifications to HT-4 (lowered atmospheric CO₂

concentration, enlarged Northern Hemisphere ice sheets, lowered sea surface temperatures). The climate values for G are indicated with an open circle connected to the climate value for HT-4 by a thin vertical line. An off-line vegetation model, forced by the seasonal cycle of temperature and precipitation, indicates a transition from forest towards grasslands in response to uplift. In southern Asia, the area of savannah/steppe is 10% (HT-2), 33% (HT-4) and 50% (G). In the Loess plateau, the area of steppe/desert is 15% (HT-2), 35% (HT-4) and 35% (G). In central Asia, the area of steppe/desert is 30% (HT-2), 70% (HT-4) and 70% (G). **c**, Wind and circulation indices for the four elevation stages and the one glacial stage: westerly jet-stream winds in December–January–February (DJF) for the western North Pacific, wind direction and relative speed (length of arrow) in JJA and DJF for the Loess plateau/east Asia region (the open arrows refer to stage G), southwesterly winds in the Arabian Sea in JJA, and an index of the intensity of the large continent-scale Asian summer monsoon, JJA, given by the sea-level pressure at the centre of the monsoon circulation.

variability of the monsoon indices but relatively small trends compared to the subsequent period. The period from about 3.6 to 2.6 Myr ago contains the most-sustained and simultaneous intensification of both summer and winter monsoons on the Loess plateau (as indicated by magnetic susceptibility, Rb/Sr, coarse grain-size fraction, and Al flux indices), as well as the most-sustained increase of aeolian flux to the North Pacific. This simultaneous intensification of both summer and winter monsoons is difficult to explain, because the rapid increase in the volume of continental ice sheets during this same period—as inferred from the marine oxygen isotope record²⁰ (Fig. 4)—implies a shift of the climate towards more glacial conditions. Based on climate-model simulations of glacial conditions, we would expect a weakening of the summer monsoon and a strengthening of the winter monsoon²¹. Therefore, we attribute the simultaneous strengthening of both summer and winter monsoons on the Loess plateau to additional, incremental plateau uplift or extension (see below). Enhanced uplift along the northern and eastern margins of the plateau after 3.6 Myr ago is inferred from widely distributed conglomerates and increased sediment flux after 3.6 Myr ago in the west Kunlun mountains²² (Fig. 4), from conglomerates dating 3.6–2.6 Myr ago in the Linxia basin²³, from a northeastward shift in maximum sedimentation rate in basins north of the east Kunlun mountains since Pliocene times²⁴, from well-developed molasse sediments as old as 3.4 Myr in the Qaidam basin²⁴, from coarse conglomerates dating from 3.6 Myr ago at Lao Junmiao on the northern flank of the Qilian mountains (Fang, X.M., personal communication), and from indications of tectonic activity (since the Pliocene) at Haiyuan, on the eastern margin of the Tibetan plateau²⁵ (Fig. 1). The magnetic susceptibility record from the Bay of Bengal also shows a rapid increase in terrigenous influx about 3.9 Myr ago (Fig. 2a). Moreover, models of plateau formation suggest continued development of the plateau towards the north and east^{26,27}.

Our climate-model simulations show that continued uplift and expansion of the plateau along its northern and eastern margins (going from stage HT-3 to HT-4, Fig. 3) enhances both summer and

winter monsoons in the region of the Loess plateau/east Asia and continues the drying trend in central Asia, but causes little change in the general Asian summer monsoon circulation or the Indian monsoon precipitation. Overall, the model results indicate that the relatively large high-elevation area that we insert in the model in going from stage HT-2 to stage HT-3, presumably reflecting elevation changes that occurred no later than about 8 Myr ago, are sufficient to alter significantly the thermally forced circulation and establish strong continent-scale summer and winter monsoons and central Asian aridity (Fig. 3). Continued elevation increases along the northern and eastern margins, in going from stage HT-3 to stage HT-4, have a more local influence restricted mainly to central Asia and the Loess plateau/east Asia sector.

The onset of major Northern Hemisphere glaciation after 2.6 Myr ago appears to have influenced, and was perhaps influenced by, the development of the Asian monsoons. After 2.6 Myr ago, the east Asian summer monsoon, reflected by the magnetic susceptibility index, becomes more variable and at times weaker (Figs 2, 4), and the phasing between orbital forcing and Indian monsoon strength changes²⁸. In contrast, the east Asian winter monsoon, reflected in the indices of the grain-size fraction and Al flux, continues strong, and even intensifies, as does the aeolian flux to the North Pacific (Figs 2, 4), indicating sustained or intensified central Asian aridity. These changes are consistent with the climate-model simulations for glacial conditions²¹: weakened summer monsoons, but continued aridity in central Asia and strong winter monsoon northwesterly winds across eastern Asia, and strong westerlies aloft (stage G, Fig. 3). The increased atmospheric dust loading associated with central Asian aridity and strong winter winds may have helped cool global climate¹¹, and thereby contributed to the development or intensification of glaciation.

We have ignored other possible influences on Late Miocene–Pliocene climates such as uplift elsewhere²⁹, changes in ocean gateways³⁰, decreases in atmospheric CO₂ concentrations due to increased weathering or carbon burial²⁹, and changes in land/ocean configuration³¹. Nevertheless, the terrestrial records from the Loess

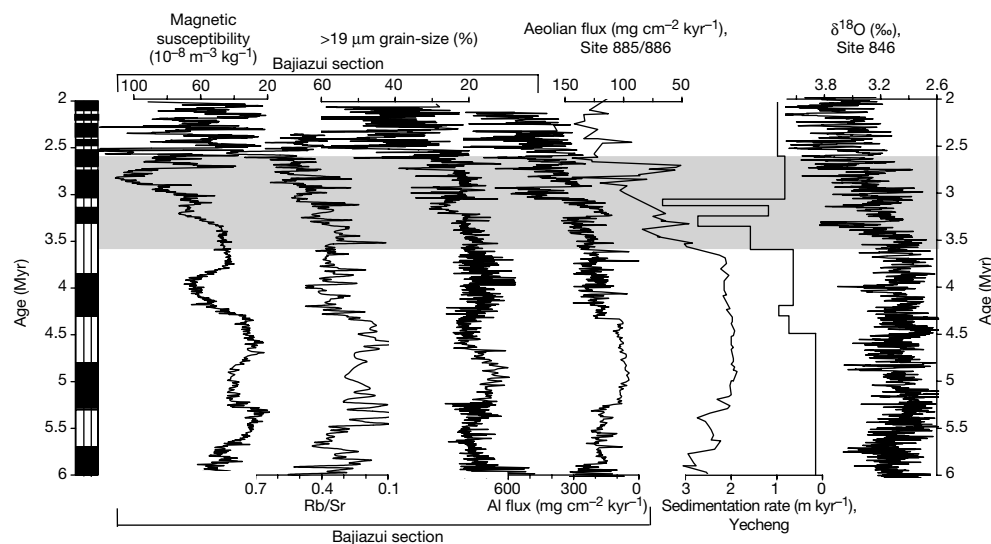


Figure 4 Terrestrial and marine records dating from 6 to about 2 Myr ago from China and the North Pacific, and indicating changes in Asian climate and global-scale glaciation. The shaded time interval between 3.6 and 2.6 Myr ago indicates the period of sustained strengthening of summer and winter monsoons on the Loess plateau. The time series are: magnetic susceptibility, Rb/Sr ratio, >19 μm grain size, and Al flux (Al content multiplied by sedimentation rate and the mean dust density of 2.5 g cm^{-3}) from the Bajiauzui section on the Loess plateau (Fig. 1); aeolian dust flux from North Pacific ODP sites 885 and 886 (Fig. 1)¹¹; sedimentation rate at Yecheng, north of the West Kunlun mountains²²; and $\delta^{18}\text{O}$ from ODP core 846 in the eastern equatorial Pacific²⁰. The chronology of monsoon proxy

sequences of the Bajiauzui section was obtained by interpolation with a sedimentation-rate model using >19 μm grain-size fraction, based on the magnetic stratigraphy⁶ and calibrated with polarity boundary ages²⁰. The palaeomagnetic results have also been confirmed by analysis of duplicate samples at the Geomagnetism Laboratory, University of Liverpool. Sampling resolution is 5–10 kyr. Sedimentation rate of the debris sequence at Yecheng is calculated based on its original thickness and polarity boundary ages. In the simplified stratigraphy column, the vertically hatched and solid black areas represent respectively loess (or loess-like) sediment and palaeosols.

plateau and the marine records from the Indian Ocean, interpreted with the aid of climate-model simulations that take into account both uplift and lateral extension of the Tibetan plateau, support and extend earlier conclusions^{11,17} concerning the nature and probable causes of the multi-stage evolution of Asian climates. □

Received 8 January; accepted 9 March 2001.

1. Kutzbach J. E., Prell, W. L. & Ruddiman W. F. Sensitivity of Eurasian climate to surface uplift of the Tibetan Plateau. *J. Geol.* **101**, 177–190 (1993).
2. Kutzbach J. E., Ruddiman, W. F. & Prell, W. L. in *Tectonic Uplift and Climate Change* (ed. Ruddiman, W. F.) 149–170 (Plenum, New York, 1997).
3. Broccoli, A. J. & Manabe, S. The effects of orography on midlatitude Northern Hemisphere dry climates. *J. Clim.* **5**, 1181–1201 (1992).
4. Molnar, P., England, P. & Martind, J. Mantle dynamics, uplift of the Tibetan Plateau and the Indian monsoon development. *Rev. Geophys.* **34**, 357–396 (1993).
5. Harrison, T. M., Copeland, P., Kidd, W. S. F. & Yin, A. Raising Tibet. *Science* **255**, 1663–1670 (1992).
6. Sun, D. H., Liu, T. S., Chen, M. Y. & Shaw, J. Magnetostratigraphy and paleoclimate of Red Clay sequences from the Chinese Loess Plateau. *Sci. China D* **40**, 337–343 (1997).
7. Sun, D. H., An, Z. S., Shaw, J., Bloemendal, J. & Sun, Y. B. Magnetostratigraphy and palaeoclimatic significance of Late Tertiary aeolian sequences in the Chinese Loess Plateau. *Geophys. J. Int.* **134**, 207–212 (1998).
8. Prell, W. L., Murray, D. W., Clemens, S. C. & Anderson, D. M. in *Synthesis of Results from Scientific Drilling in the Indian Ocean* (eds Duncan, R. A. *et al.*) 447–469 (Geophysical Monograph Series 70, American Geophysical Union, Washington DC, 1992).
9. Prell, W. L. & Kutzbach, J. E. in *Tectonic Uplift and Climate Change* (ed. Ruddiman, W. F.) 172–203 (Plenum, New York, 1997).
10. Kroon, D., Steens, T. N. F. & Troelstra, S. R. Onset of monsoonal related upwelling in the western Arabian Sea. *Proc. ODP Sci. Res.* **117**, 257–263 (1991).
11. Rea, D. K., Snoeck, H. & Joseph, L. H. Late Cenozoic eolian deposition in the North Pacific: Asian drying, Tibetan uplift, and cooling of the Northern Hemisphere. *Paleoceanography* **13**, 215–224 (1998).
12. Quade, J., Cerling, T. E. & Bowman, J. R. Development of Asian monsoon revealed by marked ecological shift in the latest Miocene of northern Pakistan. *Nature* **342**, 163–166 (1989).
13. Cerling, T. E. *et al.* Global vegetation change through the Miocene/Pliocene boundary. *Nature* **389**, 153–158 (1997).
14. Ma, Y. Z., Li, J. J. & Fang, X. M. Pollen assemblage in 30.6–5.0 Ma redbeds of Linxia region and climate evolution. *Chinese Sci. Bull.* **43**, 301–304 (1998).
15. Porter, S. C. & An, Z. S. Correlation between climate events in the North Atlantic and China during the last glaciation. *Nature* **375**, 305–308 (1995).
16. Kukla, G. & An, Z. S. Loess stratigraphy in central China. *Paleogeogr. Paleoclimatol. Paleocool.* **72**, 203–225 (1989).
17. An, Z. S. *et al.* Eolian evidence from the Chinese Loess Plateau: the onset of the late Cenozoic Great Glaciation in the Northern Hemisphere and Qinghai-Xizang Plateau uplift forcing. *Sci. China D* **42**, 258–271 (1999).
18. Maher, B. A. Characterization of soils by mineral magnetic measurements. *Phys. Earth Planet. Inter.* **42**, 76–92 (1986).
19. Chen, J., An, Z. S. & Head, J. Variation of Rb/Sr ratios in the loess-paleosol sequences of central China during the last 130,000 years and their implications for monsoon paleoclimatology. *Quat. Res.* **51**, 215–219 (1999).
20. Shackleton, N. J., Hall, M. A. & Pate, D. Pliocene stable isotope stratigraphy of site 846. *Proc. ODP Sci. Res.* **138**, 337–355 (1995).
21. Prell, W. L. & Kutzbach, J. E. Sensitivity of the Indian monsoon to forcing parameters and implications for its evolution. *Nature* **360**, 647–652 (1992).
22. Zheng, H., Powell, C., An, Z., Zhou, J. & Dong, G. Pliocene uplift of the northern Tibetan Plateau. *Geology* **28**, 715–718 (2000).
23. Li, J. J. *et al.* Late Cenozoic magnetostratigraphy (11–0 Ma) of the Dongshandian and Wangjiashan section in the Longzhong Basin, western China. *Geol. Mijnbouw* **76**, 121–134 (1997).
24. Metivier, F., Gaudemer, Y., Tapponnier, P. & Meyer, B. Northeastward growth of the Tibet plateau deduced from balanced reconstruction of two depositional areas: the Qaidam and Hexi Corridor basins, China. *Tectonics* **17**, 823–842 (1998).
25. Burchfiel, B. C. *et al.* Geology of the Haiyuan fault zone, Ningxia-Hui Autonomous region, China, and its relation to the evolution of the northeastern margin of the Tibetan Plateau. *Tectonics* **10**, 1091–1110 (1991).
26. Royden, L. H. *et al.* Surface deformation and lower crustal flow in Eastern Tibet. *Science* **276**, 788–790 (1997).
27. England, P. & Houseman, G. Finite strain calculations of continental deformation. Comparison with India-Asia collision. *J. Geophys. Res.* **91**, 3664–3676 (1986).
28. Clemens, S., Murray, D. W. & Prell, W. L. Nonstationary phase of the Plio-Pleistocene Asian monsoon. *Science* **274**, 943–948 (1996).
29. Raymo, M. E., Ruddiman, W. F. & Froelich, P. N. Influence of late Cenozoic mountain building on ocean geochemical cycles. *Geology* **16**, 649–653 (1988).
30. Haug, G. H. & Tiedemann, R. Effect of the formation of the Isthmus of Panama on Atlantic Ocean thermohaline circulation. *Nature* **393**, 673–676 (1998).
31. Ramstein, G., Fluteau, E., Besse, J. & Jossauume, S. Effect of orogeny, plate motion and land-sea distribution on Eurasian climate change over the past 30 million years. *Nature* **386**, 788–795 (1997).

Acknowledgements

The model simulations used climate models and computer resources provided by the National Center for Atmospheric Research in Boulder, Colorado. This work was supported by the Chinese Academy of Sciences, the Chinese Ministry of Science and Technology, and the Chinese National Science Foundation, and by grants to the University of Wisconsin and Brown University from the US National Science Foundation.

Correspondence and requests for materials should be addressed to J.E.K. (e-mail: jek@facstaff.wisc.edu).

Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean

Sergio A. Sañudo-Wilhelmy*, Adam B. Kustka*, Christopher J. Gobler†, David A. Hutchins‡, Min Yang*, Kamazima Lwiza*, James Burns§, Douglas G. Capone§, John A. Raven|| & Edward J. Carpenter¶

* Marine Sciences Research Center, State University of New York, Stony Brook, New York 11794-5000, USA

† Southampton College, Natural Science Division, Long Island University, Southampton, New York 11968, USA

‡ College of Marine Studies, University of Delaware, Lewes, Delaware 19958, USA

§ Wrigley Institute for Environmental Studies and Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA

|| Division of Environmental and Applied Biology, School of Life Sciences, University of Dundee, Dundee DD1 4HN, UK

¶ Romberg Tiburon Center, San Francisco State University, Tiburon, California 94920, USA

Marine fixation of atmospheric nitrogen is believed to be an important source of biologically useful nitrogen to ocean surface waters¹, stimulating productivity of phytoplankton and so influencing the global carbon cycle². The majority of nitrogen fixation in tropical waters is carried out by the marine cyanobacterium *Trichodesmium*³, which supplies more than half of the new nitrogen used for primary production⁴. Although the factors controlling marine nitrogen fixation remain poorly understood, it has been thought that nitrogen fixation is limited by iron availability in the ocean^{2,5}. This was inferred from the high iron requirement estimated for growth of nitrogen fixing organisms⁶ and the higher apparent densities of *Trichodesmium* where aeolian iron inputs are plentiful⁷. Here we report that nitrogen fixation rates in the central Atlantic appear to be independent of both dissolved iron levels in sea water and iron content in *Trichodesmium* colonies. Nitrogen fixation was, instead, highly correlated to the phosphorus content of *Trichodesmium* and was enhanced at higher irradiance. Furthermore, our calculations suggest that the structural iron requirement for the growth of nitrogen-fixing organisms is much lower than previously calculated⁶. Although iron deficiency could still potentially limit growth of nitrogen-fixing organisms in regions of low iron availability—for example, in the subtropical North Pacific Ocean—our observations suggest that marine nitrogen fixation is not solely regulated by iron supply.

We collected surface water samples and colonies of *Trichodesmium* spp. using trace-metal clean methods along two transects in the tropical (0–6° N latitude; 50–28° W longitude) and subtropical (10–16° N; 30–55° W) Atlantic Ocean in April 1996, and analysed them for C, N, P and Fe content. We also measured N₂ fixation rates of colonies (Methods).

Strong spatial gradients in the N₂-fixing diazotrophic activity were observed along the tropical and subtropical transects. Cell C specific N₂ fixation in the subtropical northern transect (median was 152 µmol N per mol C per h) was four times higher than in the tropical transect (median was 38 µmol N per mol C per h) (Fig. 1a). *Trichodesmium* biomass (Fig. 1b) was also seven times higher in the northern subtropical transect (subtropical median was 1.44 per mg chl *a* per m²; tropical median was 0.20 per mg chl *a* per m²).

In contrast to N₂ fixation, dissolved Fe concentrations in surface waters of the sub-tropical (median was 0.77 nM) and tropical (median was 0.95 nM) Atlantic were relatively constant (Fig. 1c). Similarly, levels of Fe in field-collected *Trichodesmium* colonies

ranged from 3 to 13 pmol per colony (median was 6 pmol per colony), and were not significantly different in the two transects (Fig. 1d). These Fe levels in the colonies were an order of magnitude lower than those previously reported⁵, owing to our implementation of trace-metal-clean techniques for collection and analysis. Moreover, *Trichodesmium* N₂ fixation and biomass were independent of the dissolved Fe concentrations and the Fe content of the colonies (Fig. 2a, b).

High N₂ fixation rates in the Atlantic were measured at relatively low Fe concentrations in the colonies (median was 36 μ mol Fe per mol C; range 22–72 μ mol Fe per mol C; Fig. 1d). Although the Fe requirements of *Trichodesmium* and other oceanic algae are not yet well known, the Fe:C ratios of the *Trichodesmium* colonies we measured in the central Atlantic Ocean were similar to the published Fe:C values for prokaryotic cyanobacteria (~ 20 μ mol Fe per mol C; ref. 8) growing on N (from nitrate) in cultures. Furthermore, the *Trichodesmium* Fe:C ratios we measured were nearly identical to the Fe:C uptake ratios of coastal diatoms (34–88 μ mol Fe per mol C, ref. 9), despite the fact that coastal species might be expected to have higher Fe requirements than open-ocean organisms⁸, owing to their adaptation to higher ambient Fe levels in coastal waters¹⁰. These observations suggest that the Fe requirements for diazotrophic marine cyanobacteria such as *Trichodesmium* may not be 100-fold higher than for NH₄⁺-assimilating phytoplankton, as previously calculated⁶.

We recalculated the theoretical iron use efficiency (IUE) for growth supported by N₂ fixation (Table 1) previously reported by

Raven⁶. We found that the mol Fe required (specific to the nitrogenase complex only and assuming comparable structural and bioenergetic requirements between autotrophic and heterotrophic diazotrophs) to fix 1 mol C per second via N₂ fixation ranged from 1.0 to 3.2 mol Fe, rather than 78.7 mol Fe (ref. 6). Considering this new structural estimate (Table 1) and the previously reported estimate for bioenergetic Fe demands⁶, N₂-fixing phytoplankton may have Fe requirements only 2.5–5.2 times greater than NH₄⁺-assimilating phytoplankton⁶. This range of Fe requirements is still rather higher than values calculated for phytoplankton growing on N from nitrate¹¹. However, we cannot discount any additional Fe requirement due to physiological phenomena not directly related to the structural and bioenergetic Fe demand of nitrogenase, such as variations in Fe catalyst stoichiometry¹².

Our results suggest that N₂ fixation in the subtropical Atlantic Ocean may not be an iron-limited process at this time of the year, when Fe inputs are relatively high. The apparent absence of Fe control over these biological processes is consistent with the relatively high background levels of dissolved Fe found in surface waters of the central Atlantic ocean (~ 1 nM; Fig. 1c). Those levels are between 2 and 5 times higher than the levels reported in the subtropical north Pacific (0.2–0.5 nM; ref. 13), probably owing to the high Saharan aeolian fluxes¹⁴ and tropical river inputs to the Atlantic¹⁵.

In contrast to the relatively invariant levels of dissolved Fe and Fe content in the colonies measured along the two Atlantic transects

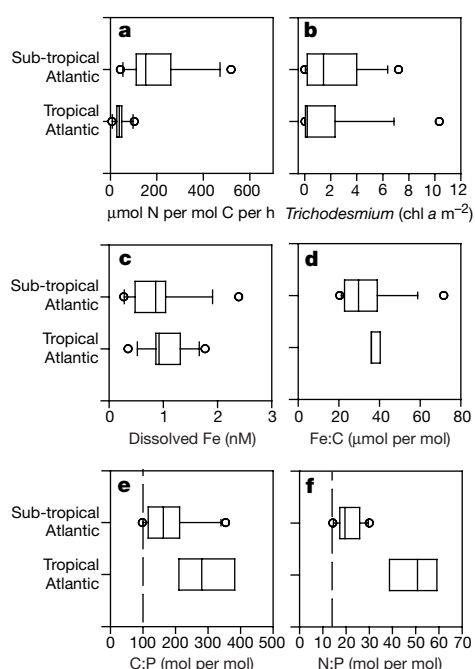


Figure 1 Box plots of N₂ fixation, dissolved Fe levels and elemental composition (C, N, P, Fe) in field-collected *Trichodesmium* colonies of the central Atlantic ocean. **a**, Nitrogen fixation. **b**, *Trichodesmium* biomass. **c**, dissolved Fe levels. **d**, Fe:C ratios. **e**, C:P ratios. **f**, N:P ratios. The tropical and subtropical areas represent our 0–6°N and 10–16°N latitude transects respectively. The line within the box is the median, and the boundary of the boxes indicates the 25th and 75th percentiles. Error bars to the left and right of the box indicate the 90th and 10th percentiles. Open circles show outlying points. Dashed line in **e** and **f** represent Redfield ratios. This figure shows that N₂ fixation (**a**) was considerably higher in the subtropical transect although the levels of dissolved Fe (**c**) in both transects were essentially the same. The Fe:C ratio (**d**) in the *Trichodesmium* colonies was relatively constant, but the C:P (**e**) and N:P (**f**) ratios were different in the subtropical and tropical transects. Those differences were caused by differences in the P content of the colonies (about two times higher in the subtropical transect). All of the data used in this figure is available in the Supplementary Information.

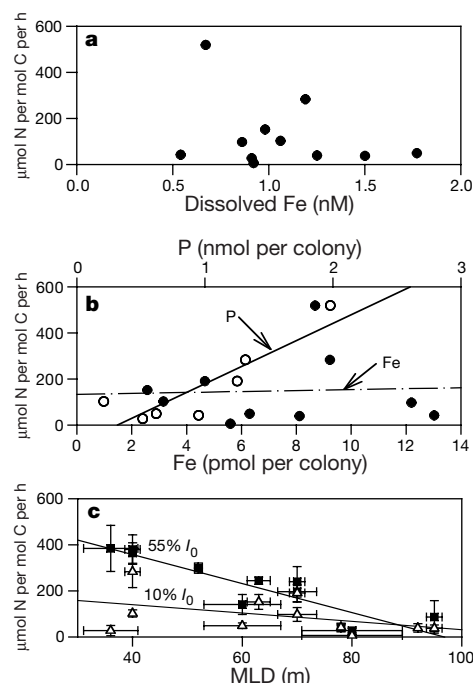


Figure 2 Relationship of N₂ fixation with dissolved Fe, with Fe and P composition in the *Trichodesmium* colonies, and with mixed-layer depth (MLD). **a**, N₂ fixation versus dissolved Fe. **b**, N₂ fixation versus Fe (filled circles) and P (open circles) content in the *Trichodesmium* colonies. Regression equation: N₂ fixation = 262 (± 65) × [P] – 83 (± 72); $r^2 = 0.77$, $P < 0.01$. **c**, N₂ fixation versus upper-ocean mixed-layer depth, calculated according to ref. 28. Squares represent N₂ fixation calculated at 55% I_0 and triangles indicate N₂ fixation calculated at 10% I_0 . I_0 , irradiance. Bars denote ± 1 s.d. (MLD, $n = 2$; N₂ fixation, $n = 3$). **a** and **b** show that N₂ fixation was independent of Fe (in the water column and in the colonies) and strongly correlated with the P content in the *Trichodesmium*. **c** shows that N₂ fixation was higher at shallower mixed-layer depth and at higher irradiance.

(Fig. 1c and d), phosphorus levels in the *Trichodesmium* colonies in the subtropical transect (median was 1.10 nmol P per colony) were two times higher than in the tropical transect (median was 0.51 nmol P per colony; Fig. 1e). Furthermore, although our analyses suggested no relationship between Fe levels and N_2 fixation in the subtropical and tropical Atlantic, rates of N_2 fixation were significantly correlated with the P levels in the colonies (Fig. 2b). Although the importance of P for N_2 fixation in the open ocean has been hypothesized^{4,16}, our data provide the first direct evidence of such a link. The N:P molar ratios measured in the colonies in the tropical (median was 51; range was 35–61) and in the subtropical (median was 21; range was 14–30; Fig. 1f) were higher than the Redfield ratio of 16:1, and these elemental ratios were within the range reported for dissolved $NO_3:PO_4$ (that is, 20–40) of the Sargasso Sea^{7,16}. Therefore, N_2 -fixing diazotrophs may contribute to the high dissolved N:P ratios observed in some areas of the Atlantic¹⁷.

Analysis of the physical regime in the central Atlantic Ocean suggested that N_2 fixation in this region could also be influenced by upper-ocean mixed-layer depth, as previously reported^{1,18,19} (Fig. 2c). The observation that N_2 fixation rates were higher at shallower mixed-layer depth is consistent with the relatively high light requirement of *Trichodesmium*²⁰, suggesting that inadequate mean irradiance for photosynthesis may directly affect the energetically expensive N_2 fixation process²¹. Our results showed that when irradiance was increased, N_2 fixation was also enhanced (Fig. 2c); therefore, light may also have limited N_2 fixation during our cruise. Surface water temperatures in the central Atlantic were relatively constant during our cruise ($26.23 \pm 1.27^\circ\text{C}$) and no relationship with N_2 fixation was observed.

Although the importance of Fe in controlling primary production in high-nitrate, low-chlorophyll regions of the world ocean is well established²², our results suggest that other factors (such as P and light) may control N_2 fixation under the high-Fe conditions of the central Atlantic ocean. Although no other measurements of Fe and P in *Trichodesmium* are available in the literature, the C:N:P:Fe stoichiometry ($99.6:18.3:1:3.7 \times 10^{-3}$) in our field-collected *Trichodesmium* colonies (Fig. 3) suggests that P should limit N_2 fixation only after the Fe quota is met. For example, an average ambient dissolved P concentration of 75 ± 42 nM in the Atlantic¹⁶ and our average dissolved Fe of 0.89 ± 0.41 nM yield a dissolved P:Fe ratio of between $1:1.1 \times 10^{-2}$ and $1:1.5 \times 10^{-2}$. Those dissolved ratios are 3–4 times lower than the ratio measured in *Trichodesmium*, suggesting that the colonies in the Atlantic were not Fe-limited during our study. However, Fe deficiency could limit the growth of this organism in other ocean basins when Fe inputs are much lower. The calculated dissolved P:Fe ratio for the subtropical north Pacific ranges from an Fe-limitation ratio of $1:9.6 \times 10^{-4}$ to a near-Fe-limitation ratio of $1:2.1 \times 10^{-3}$ (based on dissolved P and Fe concentrations of 222 ± 14 nM (ref. 16) and $0.2\text{--}0.5$ nM (ref. 13) respectively).

Our results suggest that N_2 fixation by *Trichodesmium* in the

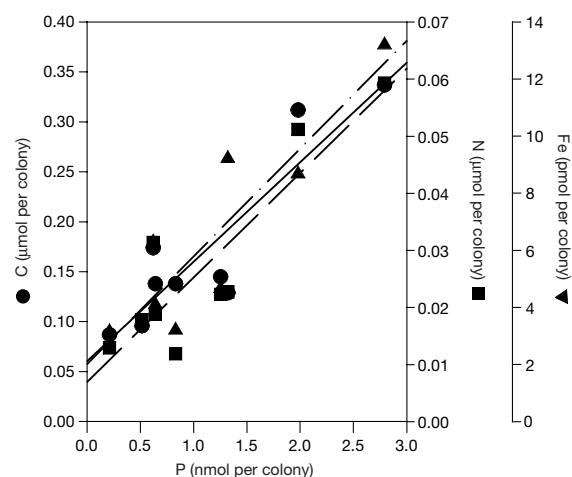


Figure 3 Elemental (C, N, P, Fe) composition of field-collected *Trichodesmium* colonies of the central Atlantic Ocean. Plot of C, N and Fe concentrations versus P concentration (denoted by circles, squares and triangles and corresponding solid, dashed and dashed-dotted regression lines, respectively). Regression equations (in mol per colony): $[C] = 99.63 (\pm 2.1) \times [P] + 6.04 \times 10^{-8} (\pm 1.4 \times 10^{-8})$, $r^2 = 0.82$. $[N] = 18.3 (\pm 0.42) \times [P] + 6.94 \times 10^{-9} (\pm 2.8 \times 10^{-9})$, $r^2 = 0.79$. $[Fe] = 3.72 \times 10^{-3} (\pm 8.7 \times 10^{-5}) \times [P] + 2.1 \times 10^{-12} (\pm 6.2 \times 10^{-13})$, $r^2 = 0.79$.

central Atlantic Ocean appears to be influenced by a series of factors (such as P or light). However, *in situ* Fe and P addition experiments at various irradiance levels are needed to further substantiate these conclusions. Rising atmospheric CO_2 may result in lower near-surface wind speeds²³ and shallower mixed layers²⁴ in the central Atlantic, consequently allowing more N_2 fixation to occur. Increased N_2 fixation could therefore potentially provide a negative feedback mechanism to climatic warming by sequestering anthropogenic CO_2 from the atmosphere. □

Methods

Surface water samples and *Trichodesmium* colonies were collected from a Zodiac deployed from the RV Seward Johnson. Water samples were filtered on-board in a portable class-100 clean unit using $0.45 \mu\text{m}$ acid-washed polycarbonate filters into acid-washed bottles. Colonies were collected at a depth of about 5 m by towing an acid-washed 102- μm plankton net at about 1 knot. In a class-100 portable bench, an average of about 100 colonies per station were hand-picked from the acid-washed cod-end using a plastic inoculating loop, deposited into 3-ml acid-washed Teflon vials, and digested in acid using a combination of Q-HCl, Q-HNO₃ and Q-HF. Fe content in the *Trichodesmium* colonies was then measured by graphite furnace atomic absorption spectrometry (GFAAS). The Fe blank was 30 ± 11 pmol per vial. Dissolved Fe was also measured using GFAAS after an organic extraction²⁵. Dissolved Fe blanks were 158 ± 88 pM (mean $\pm$ 1 s.d.). Phosphorus content in digested colonies was determined by spectrophotometric techniques developed for small volumes of interstitial waters²⁶. Nitrogen fixation was determined on isolated colonies collected between 10 and 20 m by slow (1 knot) plankton tows using a 1-m diameter, 202- μm mesh net using the C_2H_2 reduction technique²⁷. Results were converted to mol N fixed using a 3:1 ratio of C_2H_2 reduced to N_2 fixed. Particulate organic carbon

Table 1 Iron requirements for diazotrophic growth

Diazotrophs	Fe in MoFe protein ²¹ (mol mol ⁻¹)	Fe in nitrogenase complex (mol mol ⁻¹)*	Relative molecular mass of MoFe	Specific activity (nmol C_2H_4 min ⁻¹ per mg MoFe)	N fixed (per mol enzyme s ⁻¹)†	N fixed (per mol enzyme Fe s ⁻¹)‡	Fe required (per mol fixed C s ⁻¹)§
<i>Azotobacter vinelandii</i>	34–38	54–58	216–270K	1,400	3.78	0.065	2.12
<i>Azotobacter chroococcum</i>	>22	42	222K	2,000	4.93	0.117	1.22
<i>Klebsiella pneumoniae</i>	32	52	218K	2,150	5.21	0.100	1.43
<i>Clostridium pasteurianum</i>	24	44	220K	2,500	6.11	0.139	1.03
<i>Rhizobium japonicum</i>	29	49	200K	1,000	2.22	0.045	3.15

* Based on the optimal ratio of 5 mol nitrogenase reductase (Fe protein) to 1 mol nitrogenase (MoFe protein)²⁰, and 4 mol of Fe per Fe protein²¹.

† Calculated using the molecular mass, the specific activities and an ethylene production: N_2 fixation stoichiometry of 3:1.

‡ Mole Fe specifically associated with the nitrogenase complex.

§ In order to express the nitrogenase Fe requirements in terms of C fixation, a C:N fixation ratio of 7:1 is assumed as in ref. 6. Therefore, the photoautotrophic Fe requirement for C and N fixation (that is, growth) is the sum of the Fe bound in the nitrogenase complex, the bioenergetic Fe needed for C fixation and respiration to fuel nitrogenase, as well as the Fe needed for net C fixation. For the latter, we used 0.9 mol Fe (per mol C s⁻¹) to compute the estimated Fe requirement for photoautotrophic growth for photoautotrophs as $1.9\text{--}4.1$ mol Fe (per mol C s⁻¹) (see ref. 6). These enzymatic iron use efficiencies are derived from a variety of terrestrial heterotrophic diazotrophs because these data are lacking for marine photoautotrophic diazotrophs. However, the functional regions of the modelled tertiary structure of *Trichodesmium* Fe protein appear to be highly conserved with respect to a heterotrophic diazotroph such as *Azotobacter vinelandii* (ref. 30).

levels in the field-collected colonies were determined using a Carlo Erba NA1500 NCS system. For *Trichodesmium* chlorophyll biomass, the contents of whole 10-l Niskin bottles from stratified depths were gravity filtered onto 5- to 10- μ m polycarbonate filters and trichome density determined by direct microscopic enumeration using phycoerythrin epifluorescence. *Trichodesmium* trichome density was converted to chlorophyll terms by a factor derived from direct extraction and determination of chlorophyll per trichome and per colony at each station. *Trichodesmium* biomass was then integrated to the upper 50 m. Standard hydrographic parameters (temperature *T*, salinity *S* and density σ_t) were measured by CTD (conductivity–temperature–depth) at each sampling location.

Received 4 January; accepted 14 March 2001.

- Karl, D. M. *et al.* Ecosystem changes in the north Pacific subtropical gyre attributed to the 1991–92 El Niño. *Nature* **373**, 230–234 (1995).
- Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO₂ in the ocean. *Nature* **387**, 272–275 (1997).
- Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B. & Carpenter, E. J. *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276**, 1221–1229 (1997).
- Karl, D. M., Letelier, R., Tupas, L., Dore, J., Christian, J. & Hebel, D. The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* **388**, 533–538 (1997).
- Rueter, J. G., Hutchins, D. A., Smith, R. W. & Unsworth, N. L. in *Marine Pelagic Cyanobacteria: Trichodesmium and Other Diazotrophs* (eds Carpenter, E. J., Capone, D. G. & Rueter, J. G.) 289–306 (Kluwer, Norwell, 1992).
- Raven, J. A. The iron and molybdenum use efficiencies of plant growth with different energy, carbon and nitrogen sources. *New Phytol.* **109**, 279–287 (1988).
- Michaels, A. F. *et al.* Inputs, losses and transformations of nitrogen and phosphorus in the pelagic North Atlantic Ocean. *Biogeochemistry* **35**, 181–226 (1996).
- Brand, L. Minimum iron requirements of marine phytoplankton and the implications for the biogeochemical control of new production. *Limnol. Oceanogr.* **36**, 1756–1771 (1991).
- Schmidt, M. A. & Hutchins, D. A. Size-fractionated biological iron and carbon uptake along a coastal to offshore transect in the NE Pacific. *Deep Sea Res. II* **46**, 2487–2503 (1999).
- Johnson, K. S., Chavez, F. P. & Friederich, G. E. Continental-shelf sediment as a primary source of iron for coastal phytoplankton. *Nature* **398**, 697–700 (1999).
- Flynn, K. J. & Hipin, C. R. Interactions between iron, light, ammonium and nitrate: insights from the construction of a dynamic model of algal physiology. *J. Phycol.* **35**, 1171–1190 (1999).
- Raven, J. A., Evans, M. C. W. & Korb, R. E. The role of trace metals in photosynthetic electron transport in O₂-evolving organisms. *Photosynth. Res.* **60**, 111–149 (1999).
- Landing, W. M. & Bruland, K. W. The contrasting biogeochemistry of iron and manganese in the Pacific Ocean. *Geochim. Cosmochim. Acta* **51**, 29–43 (1987).
- Duce, R. A. & Tindale, N. W. Atmospheric transport of iron and its deposition in the ocean. *Limnol. Oceanogr.* **36**, 1715–1726 (1991).
- Rutgers van der Loeff, M., Helmers, E. & Kattner, G. Continuous transects of cadmium, copper, and aluminum in surface waters of the Atlantic ocean, 50° N to 50° S: Correspondence and contrast with nutrient-like behaviour. *Geochim. Cosmochim. Acta* **61**, 47–61 (1997).
- Wu, J., Sunda, W., Boyle, E. A. & Karl, D. M. Phosphate depletion in the western North Atlantic Ocean. *Science* **289**, 759–762 (2000).
- Gruber, N. & Sarmiento, J. Global patterns of marine nitrogen fixation and denitrification. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
- Mague, T. H., Mague, F. C. & Holm-Hansen, O. Physiology and chemical composition of nitrogen-fixing phytoplankton in the central north Pacific ocean. *Mar. Biol.* **41**, 213–227 (1977).
- Carpenter, E. J. & Price, C. C. Marine Oscillatoria (*Trichodesmium*) explanation for aerobic nitrogen fixation without heterocysts. *Science* **191**, 1278–1280 (1976).
- Carpenter, E. J. & Roenneberg, T. The marine planktonic cyanobacteria *Trichodesmium* spp.: photo-synthetic rate measurements in the SW Atlantic ocean. *Mar. Ecol. Prog. Ser.* **118**, 267–273 (1995).
- Postgate, J. *Nitrogen Fixation* 3rd edn 1–120 (Cambridge Univ. Press, Cambridge, 1998).
- Coale, K. H. *et al.* A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific ocean. *Nature* **383**, 495–501 (1996).
- Zwiers, F. W. & Kharin, V. V. Changes in the extremes of the climate simulated by CCC GCM2 under CO₂ doubling. *J. Clim.* **11**, 2200–2222 (1998).
- Lagerloef, G. S. E., Lukas, R., Weller, R. A. & Anderson, S. P. Pacific warm pool temperature regulation during TOGA COARE: upper ocean feedback. *J. Clim.* **11**, 2297–2309 (1998).
- Bruland, K. W., Franks, R. P., Knauer, G. A. & Martin, J. H. Sampling and analytical methods for the determination of copper, cadmium, zinc and nickel at the nanogram per liter level in sea water. *Anal. Chim. Acta* **105**, 233–245 (1979).
- Gieskes, J. M., Gamo, T. & Brumsack, H. Chemical methods for interstitial water analysis aboard Joides Resolution. Ocean Drilling Program, Technical Report No. 15 46–47 (ODP Texas A&M University, College Station, 1991).
- Capone, D. G. in *Handbook of Methods in Aquatic Microbial Ecology* (eds Kemp, P. F., Sherr, B. F., Sherr, E. B. & Cole, J. J.) 621–631 (Lewis, Boca Raton, 1993).
- Kara, A. B., Rochford, P. A. & Hurlburt, H. E. An optimal definition for ocean mixed layer depth. *J. Geophys. Res.* **105**, 16803–16821 (2000).
- Sprent, J. I. & Raven, J. A. Evolution of nitrogen-fixing symbioses. *Proc. R. Soc. Edinburgh B* **85**, 215–237 (1985).
- Zehr, J. P., Harris, D., Dominic, B. & Salerno, J. Structural analysis of the *Trichodesmium* nitrogenase iron protein: implications for aerobic nitrogen fixation activity. *FEMS Microbiol. Lett.* **153**, 303–309 (1997).

Supplementary information is available from Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by NSF Chemical and Biological Oceanography.

Correspondence and requests for materials should be addressed to S.A.S.W. (e-mail: ssanudo@notes.cc.sunysb.edu).

Preservation of ancient and fertile lithospheric mantle beneath the southwestern United States

Cin-Ty Lee^{*†}, Qingzhu Yin^{*}, Roberta L. Rudnick^{*†} & Stein B. Jacobsen^{*}

^{*} Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA

Stable continental regions, free from tectonic activity, are generally found only within ancient cratons—the centres of continents which formed in the Archaean era, 4.0–2.5 Gyr ago. But in the Cordilleran mountain belt of western North America some younger (middle Proterozoic) regions have remained stable^{1,2}, whereas some older (late Archaean) regions have been tectonically disturbed^{1,3}, suggesting that age alone does not determine lithospheric strength and crustal stability. Here we report rhenium–osmium isotope and mineral compositions of peridotite xenoliths from two regions of the Cordilleran mountain belt. We found that the younger, undeformed Colorado plateau is underlain by lithospheric mantle that is ‘depleted’ (deficient in minerals extracted by partial melting of the rock), whereas the older (Archaean), yet deformed, southern Basin and Range province is underlain by ‘fertile’ lithospheric mantle (not depleted by melt extraction). We suggest that the apparent relationship between composition and lithospheric strength, inferred from different degrees of crustal deformation, occurs because depleted mantle is intrinsically less dense than fertile mantle (due to iron having been lost when melt was extracted from the rock). This allows the depleted mantle to form a thicker thermal boundary layer⁴ between the deep convecting mantle and the crust, thus reducing tectonic activity at the surface. The inference that not all Archaean crust developed a strong and thick thermal boundary layer leads to the possibility that such ancient crust may have been overlooked because of its intensive reworking or lost from the geological record owing to preferential recycling.

The North American Cordillera is a broad continental region marked by a long period of tectonic activity, which began in the Palaeozoic era with a series of mountain-forming events and culminated in the Cenozoic era with extension⁵. Deformation appears to be heterogeneously distributed (Fig. 1). The Basin and Range province, which includes much of Nevada and southeastern California, experienced crustal thickening and subsequent large-scale extension (possibly up to 200%)⁶. In contrast, the Colorado plateau, an elevated circular region surrounded on all sides by deformed crust, has remained an island of tectonic quiescence, as evidenced by flat-lying, unfolded and largely unfaulted Palaeozoic sedimentary strata⁷.

Given the relative differences in the degree of deformation seen in the overlying crust⁵ and the correlation between age and stability observed elsewhere in the continents, the more-tectonized Basin and Range lithosphere might be expected to be younger than that beneath the less-tectonized Colorado plateau. However, Sm–Nd model ages indicate that the crust in the southern Basin and Range (referred to here as Mojavia) is older, formed in Palaeoproterozoic to Archaean times (~2.0–2.6 Gyr ago)^{1,2}, whereas the Colorado plateau crust formed subsequently in the middle Proterozoic (1.6–2.0 Gyr ago)^{1,3}. There are two possible explanations for this unexpected relationship. First, the lithospheric mantle beneath Mojavia may not be as old as the crustal model ages indicate. This might

[†]Present address: Division of Geological and Planetary Sciences, California Institute of Technology, 1200 E. California Boulevard, Pasadena, California 91125, USA (C.-T.L.); Department of Geology, University of Maryland, College Park, Maryland 20742, USA (R.L.R.).

occur if the Nd model ages reflect mixing between juvenile middle Proterozoic crust and sedimentary material shed from an adjacent Archaean craton at the time of lithosphere formation¹, or the original ancient lithospheric mantle was removed (as recently suggested for the deep lithosphere of the Sierra Nevada⁸). Second, the lithospheric mantle beneath both regions may be the same age as

the overlying crust⁹, but the ancient lithospheric mantle beneath Mojavia is inherently weaker than typical cratonic lithospheric mantle owing to a more fertile composition^{4,10} and/or the presence of water¹¹. We evaluate these two alternatives using Os model ages of peridotite xenoliths recently derived from the lithospheric mantle in both regions.

Assuming that partial melting leads to stabilization of the lithospheric mantle, the Re–Os isotope systematics of peridotite xenoliths (samples of the lithospheric mantle) can be used to date this time of stabilization; this is because partial melting fractionates Re/Os (Re is moderately depleted and Os is sequestered in the residue¹²). As ¹⁸⁷Re decays to ¹⁸⁷Os (half life of ~42 Gyr), the residue evolves towards unradiogenic ¹⁸⁷Os/¹⁸⁸Os and diverges from the isotopic trajectory of undepleted convecting mantle¹². Analogous to Sm–Nd model ages, the time at which melt extraction occurred is determined from the intersection of the two isotopic evolution paths, as constrained by the present-day ¹⁸⁷Os/¹⁸⁸Os and ¹⁸⁷Re/¹⁸⁸Os ratios of the sample and the convecting mantle. Because the Os concentration in peridotites is 10–1,000 times greater than that of silicate melts or other metasomatic agents, the Os isotope system in peridotites is considerably more robust to changes imparted by metasomatism and contamination than are isotope systems based on incompatible elements, such as Sm–Nd and Rb–Sr.

The peridotite xenoliths come from the Pliocene Cima volcanic field, located in the Mojavia province, California, and from Miocene–Oligocene minette plugs located in the ‘Four corners’ region of the Colorado plateau. The Cima peridotites are fresh and devoid of hydrous minerals, while the Colorado plateau samples are partly serpentinized. The Mojavian samples have ¹⁸⁷Os/¹⁸⁸Os ratios ranging from chondritic to relatively unradiogenic values (see Supplementary Information). The latter reflect long-term isolation in a low Re/Os environment produced by ancient partial melting. The samples show a positive correlation on an isochron plot, with a slope corresponding to an age of 2.5 ± 1.6 Gyr and an initial ¹⁸⁷Os/¹⁸⁸Os of ~0.112 (Fig. 2a). Using Al₂O₃ as a proxy for the Re/Os ratio¹³ results in an initial ¹⁸⁷Os/¹⁸⁸Os of ~0.113 (Fig. 2b), indicating that depletion in Re and Al probably reflect the same ancient partial melting event. The oldest Re depletion age (assuming Re/Os = 0) for the Mojave samples is 2.4 Gyr (¹⁸⁷Os/¹⁸⁸Os = 0.1120). Although this represents a minimum age, this sample’s ¹⁸⁷Os/¹⁸⁸Os may closely approximate the ‘true’ initial ratio because its Al₂O₃ content (0.68 wt%) is below the lower limit (0.7 wt%) at which

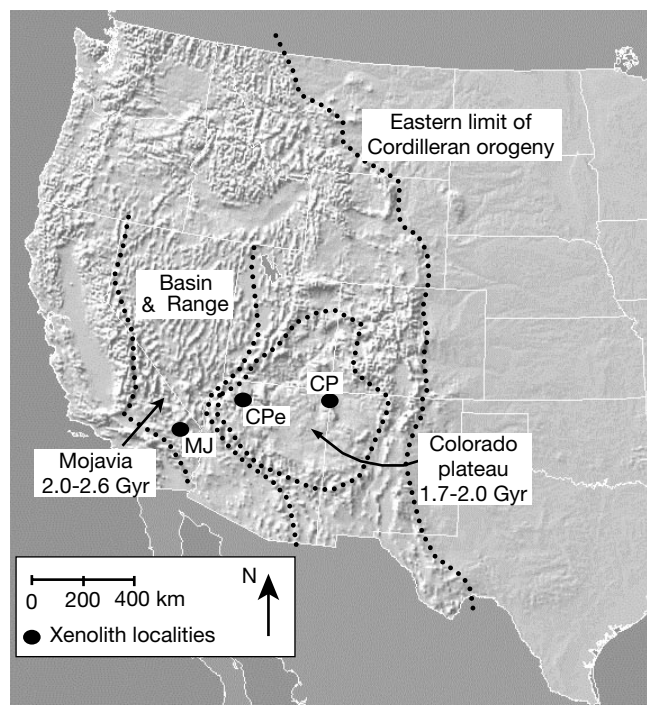


Figure 1 Digital elevation map showing topographic features within the Cordilleran orogenic belt, most of which are associated with late Cenozoic tectonism. The Colorado plateau is an elevated region of mostly undeformed sedimentary strata, surrounded on all sides by regions that have undergone compression and/or extension. Ages represent Sm–Nd crust formation ages^{1,2}. Xenolith localities are indicated: CP, central Colorado plateau (The Thumb volcanic dyke); CPe, edge of Colorado plateau (Grand Canyon volcanic field); and MJ, Mojavia (Cima volcanic field).

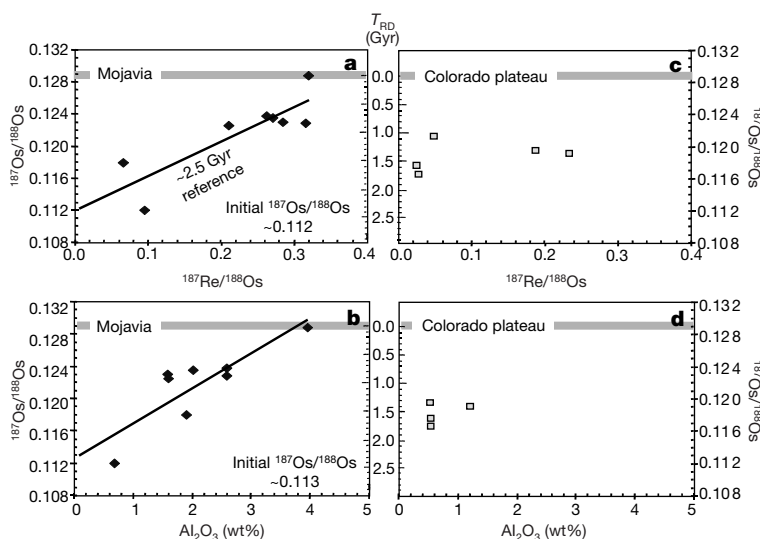


Figure 2 Re–Os isotopic constraints on the age of lithospheric mantle beneath Mojavia and the Colorado plateau. Top panels, Re–Os isotope data (a, Mojavia; c, Colorado plateau). Bottom panels, Al₂O₃ content (wt%) used as a proxy for Re/Os (b, Mojavia;

d, Colorado plateau). T_{RD} represents Re-depletion model ages, calculated by assuming Re/Os = 0 and extraction from a primitive upper mantle¹² (grey horizontal bar).

Re/Os is suggested to drop to zero during partial melting of the upper mantle¹⁴. Re–Os model ages on individual samples range between 1.8 and 3.4 Gyr (except for Ki5-139). Collectively, these data point to a late Archaean or earliest Proterozoic age for lithospheric mantle formation beneath Mojavia. This age is consistent with the Sm–Nd model ages of the overlying crust, which thus appear to represent accurately the timing of crust formation^{1,2,9}.

The Colorado plateau samples (see Supplementary Information) plot along a horizontal array on a Re–Os isochron diagram (Fig. 2c), which does not give any meaningful age significance. Instead, the array is probably due to recent addition or loss of Re (ref. 12). Although the data also do not define a strong correlation between $^{187}\text{Os}/^{188}\text{Os}$ and Al_2O_3 content (Fig. 2d), the low Al_2O_3 contents and $^{187}\text{Os}/^{188}\text{Os}$ imply that most of the Re was depleted during a partial melting event, which occurred ~ 1.6 Gyr ago (Fig. 2c, d), within uncertainty of Sm–Nd model ages of eclogitic xenoliths from the same region³ and consistent with evidence for old Nd in the lithospheric mantle beneath the Colorado plateau^{15,16}.

Our data not only confirm that ancient lithospheric mantle persists beneath the southwestern USA despite the protracted Cordilleran orogeny, but also show that the age of the lithospheric

mantle beneath each province is indistinguishable from Nd model ages of the overlying crust. Thus continental lithospheric strength is not strictly a function of the formation age of the lithospheric mantle, as might be inferred from the stability of Archaean cratons.

Instead, bulk composition may control the strength of continental lithospheric mantle, by controlling its ultimate thickness. To quantify composition, we use the Mg number ($\text{Mg\#}$ is the molar $\text{Mg}/(\text{Mg}+\text{Fe})$ ratio) of a peridotite as a measure of the degree of depletion. This number is an inverse measure of the amount of Fe, and the higher the $\text{Mg\#}$, the lower the density. The mode $\text{Mg\#}$ for the lithospheric mantle of strong Archaean cratons, such as Tanzania^{17,18}, South Africa^{19,20} and Siberia²¹, is 0.93, for the weak Palaeoproterozoic to Archaean Mojavia province it is 0.90 (based on compilations in ref. 22 and our own data), and for the strong Proterozoic Colorado plateau it is 0.91 (refs 15, 23) (Fig. 3). Previous studies have shown that most Archaean cratons are underlain by more depleted mantle than post-Archaean regions, leading to an apparent correlation between strength and bulk composition. Our work corroborates these findings by showing that even when age–strength correlations break down, bulk composition still influences lithospheric strength.

Motivated by seismological evidence and the lack of a strong correlation between continents and the long-wavelength geoid, Jordan suggested that continents are underlain by thermal boundary layers, stabilized against convective disruption by compositional buoyancy⁴. This condition, known as the ‘isopycnal hypothesis’, requires that the negative buoyancy imposed by the colder thermal state of the mantle beneath continents is exactly compensated by a lower intrinsic density of the mantle beneath continents at every level within the viscous thermal boundary layer (but below the elastic mechanical boundary layer). Figure 4 shows isopycnal density curves calculated (see Methods) for various ocean–continent

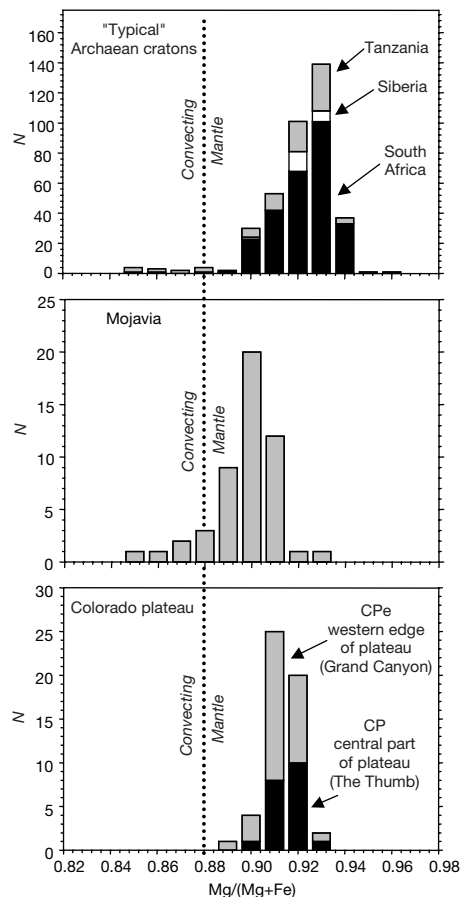


Figure 3 Comparison of the bulk $\text{Mg\#}$ of lithospheric mantle beneath typical Archaean cratons, the tectonized Mojavia block, and the strong Colorado plateau, N , number of samples. Top, a compilation of bulk-rock $\text{Mg\#}$ for low-temperature peridotites from the Tanzanian^{17,18}, Siberian²¹ and South African cratons^{19,20}. High-temperature sheared peridotites were not included in the compilation because these may have been recently refertilized by magmatic processes shortly before eruption. Middle, data from Mojavian xenolith localities based on bulk-rock compilations from ref. 22 and our own data set (olivine compositions). Bottom, data from Colorado plateau samples, which include peridotites from the western edge of the plateau (Grand Canyon¹⁵) and from central part of the plateau (the ‘Four corners’ region^{10,23}). The estimated $\text{Mg\#}$ for the convecting upper mantle is taken to be ~ 0.88 (shown as a vertical dotted line).

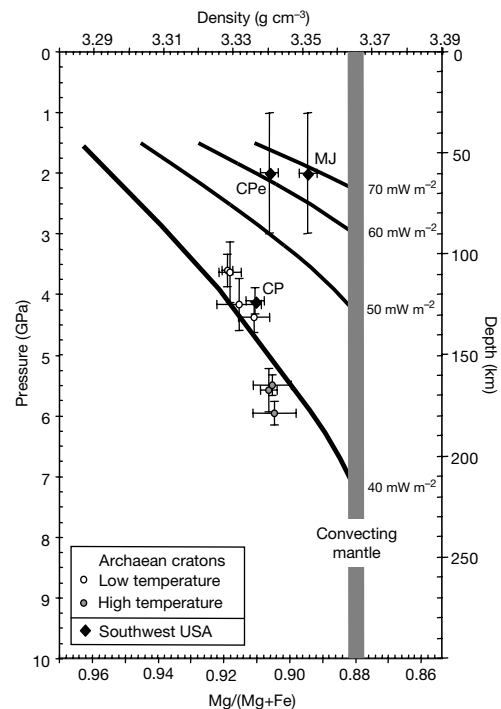


Figure 4 Isopycnal density curves, illustrating the role of bulk composition in determining lithospheric thickness. Curves were calculated for a range of surface heat flows: 70, 60, 50 and 40 mW m^{-2} . MJ, CP and CPe are defined in Fig. 1 legend. Unlabelled points refer to data from different xenolith suites from the Siberian, Tanzanian, and South African cratons (high-temperature peridotites are also shown for reference). Each symbol represents the average density and pressures of all xenoliths within a given xenolith suite. See Methods for details.

temperature differences for a range of surface heat flows and parameters given in the figure legend²⁴ and in Methods. Although the absolute positions of the isopycnal lines depend strongly on the assumed parameters (namely, crustal heat production and potential temperature of the convecting mantle adiabat), Fig. 4 illustrates that relative differences in bulk composition translate into differences in the thickness of the thermal boundary layer regardless of any uncertainty in the above parameters. The isopycnal hypothesis thus explains the observed correlation between strength and bulk composition because more depleted mantle generates a thicker, colder and hence stronger thermal boundary layer, whereas more-fertile mantle allows for a thinner, hotter and hence weaker thermal boundary layer.

Because the Mojavian peridotites are more fertile than typical cratonic mantle and derive from shallower depths (based on the lack of garnet), they plot on a warmer isopycnal curve ($\sim 70 \text{ mW m}^{-2}$), whereas the xenoliths from the centre of the Colorado plateau (garnet-bearing) plot near the array defined by peridotites from stable Archaean cratons. Ignoring localized regions of high heat flow attributed to recent magmatism, we note that the measured surface heat flow for Mojavia is $60\text{--}80 \text{ mW m}^{-2}$ and that for the Colorado plateau ranges to lower values ($40\text{--}80 \text{ mW m}^{-2}$)²⁵. If all of these regions originally fit the isopycnal condition, as is probably the case, given the lack of large negative geoid or free air gravity anomalies over continents²⁶, the more-fertile character of the Mojavia mantle requires a thinner thermal boundary layer ($50\text{--}100 \text{ km}$) than that beneath the more-depleted Colorado plateau and stable Archaean cratons ($\sim 200 \text{ km}$). This is consistent with the lack of garnet in Mojavian xenoliths (implying lithospheric thicknesses less than $\sim 90 \text{ km}$) and the presence of garnet in the Colorado plateau xenoliths. Thermobarometric data on the latter (calculated from the data of ref. 23) show that the Colorado plateau lithosphere extended to at least 120 km depth at the time of xenolith sampling. In this context, we note that spinel peridotite xenoliths from the western edge of the Colorado plateau^{10,15} plot along an isopycnal line intermediate between Mojavia and the central Colorado plateau, suggesting that the Colorado plateau lithospheric mantle is thickest at its centre.

The correlation between strength and bulk composition can be explained if the latter dictates the ultimate thickness to which the thermal boundary layer can grow beneath continents, in turn controlling lithospheric strength by defining the thickness of the lithosphere itself. Further thickening of the thermal boundary layer beyond that allowed by the bulk Mg# of the lithospheric mantle would result in this mantle becoming unstable due to its increased negative thermal buoyancy. We suggest that the low viscosity imposed on the residual peridotite by the high temperatures associated with the melting event may have been compensated by a viscosity increase produced by complete dehydration, thus explaining how originally hot, depleted mantle may have eventually stabilized to form thick lithospheric mantle^{11,27}.

Our findings also have implications for interpreting the ancient rock record. As the Earth may have been hotter and convecting more vigorously in the past, there may have been a secular decrease in the degree of melting, resulting in an apparent correlation between lithospheric mantle composition, and hence lithospheric strength, with age. However, the unusually fertile, ancient lithospheric mantle in the Mojavia province raises the question of how much Archaean lithosphere, characterized by only minor depletion, may have been lost from the geologic record by recycling back into the mantle. If this has occurred, the present distribution of Archaean crust may thus represent a biased sampling of crust underlain by highly depleted mantle (as suggested, for example, in ref. 28). Alternatively, the fact that Mojavian lithospheric mantle still persists despite its fertility may mean that even minor depletion is sufficient to inhibit recycling of continental lithosphere. However, because of its weaker character, such fertile, ancient lithosphere may be more easily

overlooked and mistaken for Phanerozoic material on the basis of the degree of deformation or metamorphic age. □

Methods

Calculation of isopycnal density curves

These curves (Fig. 4) were calculated from $\Delta\rho(z) = \rho^0\alpha\Delta T(z)$, where ρ^0 is the normative density (at STP) of the fertile convecting mantle beneath oceans (grey vertical bar), α is the thermal expansion ($2.7 \times 10^{-5} \text{ per } ^\circ\text{C}$), and $\Delta T(z)$ is the continent–ocean temperature difference as a function of depth z (ref. 4). Values of $\Delta T(z)$ were determined by assuming a $1,300^\circ\text{C}$ potential adiabat for the convecting mantle (adiabatic temperature gradient of $0.5^\circ\text{C km}^{-1}$), and continental geotherms with a range of surface heat flows; we assume that heat production of the crust and mantle is $0.5 \text{ } \mu\text{W m}^{-3}$ and $0.02 \text{ } \mu\text{W m}^{-3}$ (refs 24, 25), respectively. This crustal heat-production value was chosen so that the average pressures and densities of peridotite xenoliths from the stable Archaean Tanzania, Siberian and South African cratons approximately coincide with an isopycnal curve derived for a surface heat flow of 40 mW m^{-2} , the average measured on stable Archaean cratons²⁹. This choice of crustal heat production is within the range allowed by xenolith thermobarometry²⁴. Densities were calculated using an empirical correlation between bulk Mg# and density (in g cm^{-3}), $\rho = 4.201 - 0.950\text{Mg\#}$ ($r^2 = 0.74$) from a suite of well-characterized xenoliths from Tanzania¹⁸. Error bars in Fig. 4 represent the $2\sigma_{\text{mean}}$ of the entire population of densities (obtained by propagating the $2\sigma_{\text{mean}}$ of Mg#s through the above equation) and pressures for a given xenolith suite (errors in the thermobarometric equation or the equation for calculating density from Mg# are not included). Tanzanian densities were calculated using a linear combination of mineral endmember compositions³⁰ and their modal abundances (calculated by least-squares regression of whole-rock and mineral compositions). For internal consistency, the density for convecting upper mantle³¹ was calculated in the same manner. Pressures for garnet-bearing peridotites (most cratonic peridotites) were calculated using thermobarometers based on the solubility of Al in orthopyroxene coexisting with garnet³². For spinel peridotites (for example, MJ and CPe), the pressure was estimated to be between 1 and 3 GPa, based on limits placed on the thickness of the crust and the absence of garnet.

Received 1 September 2000; accepted 27 February 2001.

- Bennett, V. C. & DePaolo, D. J. Proterozoic crustal history of the western United States as determined by neodymium isotopic mapping. *Geol. Soc. Am. Bull.* **99**, 674–685 (1987).
- Ramo, O. T. & Calzia, J. P. Nd isotopic composition of cratonic rocks in the southern Death Valley region: evidence for a substantial Archaean source component in Mojavia. *Geology* **26**, 891–894 (1998).
- Wendlandt, E., DePaolo, D. J. & Baldrige, W. S. Nd and Sr isotope geochronology of Colorado Plateau lithosphere: implications for magmatic and tectonic underplating of the continental crust. *Earth Planet. Sci. Lett.* **116**, 23–43 (1993).
- Jordan, T. H. Composition and development of the continental tectosphere. *Nature* **274**, 544–548 (1978).
- Burchfiel, B. C. & Davis, G. A. Nature and controls of Cordilleran orogenesis, western United States: extensions of an earlier synthesis. *Am. J. Sci.* **275**, 363–396 (1975).
- Wernicke, B., Axen, G. J. & Snow, J. K. Basin and Range extensional tectonics at the latitude of Las Vegas, Nevada. *Geol. Soc. Am. Bull.* **100**, 1738–1757 (1988).
- Dickinson, W. R. in *Geologic Evolution of Arizona* (eds Jenney, J. P. & Reynolds, S. J.) 1–16 (Arizona Geological Society Digest, Vol. 17, Arizona Geological Survey, Tucson, AZ, 1989).
- Lee, C.-T., Yin, Q.-Z., Rudnick, R. L., Chesley, J. T. & Jacobsen, S. B. Os isotopic evidence for Mesozoic removal of lithospheric mantle beneath the Sierra Nevada, California. *Science* **289**, 1912–1916 (2000).
- Wooden, J. L. & Miller, D. M. Chronologic and isotopic framework for early Proterozoic crustal evolution in the eastern Mojave desert region, SE California. *J. Geophys. Res.* **95**, 20133–20146 (1990).
- Smith, D. Insights into the evolution of the uppermost continental mantle from xenolith localities on and near the Colorado Plateau and regional comparisons. *J. Geophys. Res.* **105**, 16769–16781 (2000).
- Pollack, H. N. Cratonization and thermal evolution of the mantle. *Earth Planet. Sci. Lett.* **80**, 175–182 (1986).
- Walker, R. J., Carlson, R. W., Shirey, S. B. & Boyd, F. R. Os, Sr, Nd, and Pb isotope systematics of southern African peridotite xenoliths: implications for the chemical evolution of subcontinental mantle. *Geochim. Cosmochim. Acta* **53**, 1583–1595 (1989).
- Reisberg, L. & Lorand, J.-P. Longevity of sub-continental mantle lithosphere from osmium isotope systematics in orogenic peridotite massifs. *Nature* **376**, 159–162 (1995).
- Handler, M. R. & Bennett, V. C. Behaviour of platinum-group elements in the subcontinental mantle of eastern Australia during variable metasomatism and melt depletion. *Geochim. Cosmochim. Acta* **63**, 3597–3618 (1999).
- Alibert, C. Peridotite xenoliths from western Grand Canyon and the Thumb: a probe into the subcontinental mantle of the Colorado Plateau. *J. Geophys. Res.* **99**, 21605–21620 (1994).
- Roden, M. F., Smith, D. & Murthy, V. R. Chemical constraints on lithosphere composition and evolution beneath the Colorado Plateau. *J. Geophys. Res.* **95**, 2811–2831 (1990).
- Rudnick, R. L., McDonough, W. F. & Orpin, A. in *Proc. Vth Int. Kimberlite Conf.* Vol. 1 (eds Meyer, H. O. A. & Leonards, O.) 336–353 (Companhia de Pesquisa de Recursos Minerais Spec. Publ. 1, Rio de Janeiro, 1994).
- Lee, C.-T. & Rudnick, R. L. in *Proc. VIIth Int. Kimberlite Conf.* P. H. Nixon Vol. (eds Gurney, J. J., Gurney, J. L., Pascoe, M. D. & Richardson, S. H.) 503–521 (Red Roof Design, Cape Town, 1999).
- Boyd, F. R. & Mertzman, S. A. in *Magmatic Processes: Physicochemical Principles* (ed. Mysen, B. O.) 13–24 (Spec. Publ. Vol. 1, Geochemical Society, University Park, Pennsylvania, 1987).
- Winterburn, P. A., Harte, B. & Gurney, J. J. Peridotite xenoliths from the Jagersfontein kimberlite pipe: I. Primary and primary-metasomatic mineralogy. *Geochim. Cosmochim. Acta* **54**, 329–341 (1990).
- Boyd, F. R. *et al.* Composition of the Siberian cratonic mantle: evidence from Udachnaya peridotite xenoliths. *Contrib. Mineral. Petrol.* **128**, 228–246 (1997).

22. Wilshire, H. G., McGuire, A. V., Noller, J. S. & Turin, B. D. Petrology of lower crustal and upper mantle xenoliths from the Cima volcanic field, California. *J. Petrol.* **32**, 169–200 (1991).
23. Ehrenberg, S. N. Petrogenesis of garnet lherzolite and megacrystalline nodules from the Thumb, Navajo Volcanic field. *J. Petrol.* **23**, 507–547 (1982).
24. Rudnick, R. L., McDonough, W. F. & O'Connell, R. J. Thermal structure, thickness and composition of continental lithosphere. *Chem. Geol.* **145**, 395–411 (1998).
25. Sass, J. H. *et al.* Thermal regime of the southern Basin and Range Province: 1. Heat flow data from Arizona and the Mojave Desert of California and Nevada. *J. Geophys. Res.* **99**, 22093–22119 (1994).
26. Shapiro, S. S., Hager, B. H. & Jordan, T. H. The continental tectosphere and Earth's long-wavelength gravity field. *Lithos* **48**, 135–152 (1999).
27. Hirth, G., Evans, R. L. & Chave, A. D. Comparison of continental and oceanic mantle electrical conductivity: is the Archean lithosphere dry? *Geochem. Geophys. Geosys.* [online] (<http://146.201.254.53/publicationsfinal/articles/2000GC000048/article2000GC000048.pdf>) (2000).
28. Bowring, S. A. & Housh, T. The Earth's early evolution. *Science* **269**, 1535–1540 (1995).
29. Nyblade, A. A. & Pollack, H. N. A global analysis of heat flow from Precambrian terranes: implications for thermal structure of Archean and Proterozoic lithosphere. *J. Geophys. Res.* **98**, 12207–12218 (1993).
30. Bass, J. D. in *Mineral Physics and Crystallography: A Handbook of Physical Constants* (ed. Ahrens, T. J.) 45–63 (American Geophysical Union, Washington, District of Columbia, 1995).
31. McDonough, W. F. The composition of the Earth. *Chem. Geol.* **120**, 1535–1540 (1995).
32. Brey, G. P. & Kohler, T. Geothermobarometry in four-phase lherzolites II. New thermobarometers, and practical assessment of existing thermobarometers. *J. Petrol.* **31**, 1353–1378 (1990).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

All samples, except for CIP98-8 and CIP98-64, were borrowed from the Smithsonian Institution (we thank S. Sorensen for her curation of these specimens). We thank J. Chesley, M. Handler, A. Nyblade, D.G. Pearson, D. Smith and R.J. Walker for comments on the manuscript. This work was supported by the NSF, and C.-T.L. was supported in part by an NSF graduate fellowship and a Mineralogical Society of America grant for research.

Correspondence and requests for materials should be addressed to C.-T.L. (e-mail: ctlee@gps.caltech.edu).

Consistent patterns and the idiosyncratic effects of biodiversity in marine ecosystems

Mark C. Emmerson⁺, Martin Solan⁺, Chas Emes⁺, David M. Paterson[†] & Dave Raffaelli^{*,‡}

^{*} Culterty Field Station, University of Aberdeen, Newburgh, Ellon, AB41 6AA, Scotland

[†] Gatty Marine Laboratory, University of St Andrews, St Andrews, Fife, KY16 8LB, Scotland

Revealing the consequences of species extinctions for ecosystem function has been a chief research goal^{1–7} and has been accompanied by enthusiastic debate^{8–11}. Studies carried out predominantly in terrestrial grassland and soil ecosystems have demonstrated that as the number of species in assembled communities increases, so too do certain ecosystem processes, such as productivity, whereas others such as decomposition can remain unaffected¹². Diversity can influence aspects of ecosystem function, but questions remain as to how generic the patterns observed are, and whether they are the product of diversity, as such, or of the functional roles and traits that characterize species in ecological systems. Here we demonstrate variable diversity effects for species representative of marine coastal systems at both global and regional scales. We provide evidence for an increase in complementary resource use as diversity increases and show strong evidence for diversity effects in naturally assembled com-

munities at a regional scale. The variability among individual species responses is consistent with a positive but idiosyncratic pattern of ecosystem function with increased diversity.

We investigated diversity–function relations in mesocosms containing a gradient of species richness, established using intertidal invertebrates. The ecosystem function that we measured was the flux of nutrients (specifically ammonia nitrogen (NH₄-N)) to the overlying water column, a microbial process essential for primary production, which is mediated through the physical working of the sediment by invertebrates^{13,14}.

Replicate species pools were sampled from locations around the coastlines of northeast Scotland, southwest Sweden and central south Australia. These replicate regional species pools allow us to separate the effects of species richness from any combinatorial effects that might result from species identity. Such between-site comparisons have been made in terrestrial systems⁴, and the importance of environment and site effects is well recognized^{15–17}.

For each site we created a species richness gradient, on the basis of previous knowledge of the local species pools. Species were not selected randomly; rather, they were common species dominating the biomass at each site, and all are known to interact directly with the sediment in which they live. These simple assemblages ranged from one to four species, dependent on locality (see Methods and Supplementary Information). At these low levels of species richness, the effects of diversity are most likely to be manifest. Additionally, species were classified into one of three functional groups so that the effects of functional diversity could also be considered (see Methods). Two treatments were employed: a primary treatment of species

Table 1 Summary of ANCOVA of NH₄-N production across sites and treatments

Source of variation	d.f.	s.s.	m.s.	F	P
Combined site analysis*					
Site	3	38.108	1.723	158.71	0.001†
Biomass	1	0.937	0.346	31.93	0.001
Site*biomass	3	0.156	0.037	3.44	0.018
Species richness	1	0.058	0.05	4.65	0.032
Error	195	2.12	0.01		
Total	211	43.82			
Individual site analysis‡					
Gullmarsfjord					
Biomass	1	0.234	0.162	50.18	0.001
Species richness	1	0.055	0.017	5.31	0.024
Function richness	1	0.109	0.014	4.49	0.037
Species identity	9	0.206	0.023	7.12	0.001
Error	77	0.248	0.003		
Total	89	0.854			
Ythan a§					
Biomass	1	0.058	0.026	5.18	0.035
Species identity	3	0.067	0.022	4.4	0.016
Error	19	0.096	0.005		
Total	25	0.233			
Ythan b§					
Biomass	1	0.301	0.098	34.09	0.001
Species identity	3	0.438	0.1463	50.92	0.001
Error	40	0.114	0.002		
Total	46	0.917			
Boston bay					
Biomass	1	0.828	0.486	25	0.001
Error	39	0.758	0.019		
Total	48	1.732			

d.f., degrees of freedom. s.s., sum of squares. m.s., mean squares.

* Each site was treated as a replicate for the purposes of a large combined analysis. the design treated site as a factor whereas biomass, species richness and functional group richness were treated as covariates. Functional group richness terms were non-significant and are not presented here.

† Significant inter-site differences are responsible for significant species richness terms (two and four species treatments are only present at Gullmarsfjord site, which is significantly different from Ythan and Boston bay sites).

‡ Because of the significant site terms each site was analysed separately. The structure of this analysis regarded species identity (treatment identity) as the main factor, and biomass, species richness and functional group richness as covariates for the purposes of regression analysis.

§ At the Ythan estuary sites, both species richness and functional richness were non-significant (data not shown).

|| No significant effects were found at Boston bay for species richness, functional richness or species identity.

⁺ Present address: Environment Department, University of York, Heslington, York, YO10 5DD, UK.

richness and a secondary treatment of increasing biomass (seven biomass levels), which was fixed across each species richness treatment (see Fig. 1 for the biomass ranges used) to define the relationship between biomass of component species and the flux of nutrients. All species were characterized in monoculture and mixtures consisted of these characterized species. Species richness treatments were replicated across the different species pools to minimize pseudoreplication. Single species treatments allow our design to identify the 'sampling effect', that is, the chance inclusion of a species with dominating traits, and allows for the determination of 'complementarity' (the complementary use of resources, considered by some to be a diversity effect^{3,18–20}). Finally, intact naturally assembled communities, which represented a greater range of species richness, were sampled from three sites around the northeast coast of Scotland, and returned to the laboratory where they were placed in similar mesocosms. We established a total of 241 mesocosms containing different species combinations and biomass levels.

In an overall analysis (see Methods) two general patterns emerge from these experiments: first, $\text{NH}_4\text{-N}$ concentration (ecosystem function) increases with invertebrate biomass (analysis of covariance: $F_{1,195} = 31.93$, $P < 0.001$ (Table 1)), (Fig. 1) and second, the responses from individual species treatments are different (dependent on species identity) and variable. We propose that these inherent differences will lead to an idiosyncratic relationship between diversity and ecosystem function as species are added sequentially to an ecosystem. Despite such inherent variability we would still expect consistent patterns (positive, negative or neutral trends) to emerge for the function in question, irrespective of species identity—the two (idiosyncratic behaviour and consistent

effects) are not mutually exclusive.

There are also clear between-site differences in the absolute levels of $\text{NH}_4\text{-N}$ ($F_{3,195} = 158.71$, $P < 0.001$ (Table 1)). Concentrations at the Swedish site were in the range $0.97\text{--}2.87\text{ mg l}^{-1}$ but at the other sites concentrations were much higher ($6.93\text{--}37.1\text{ mg l}^{-1}$). In the overall analysis there was also a significant effect of species richness ($F_{1,195} = 4.65$, $P < 0.032$ (Table 1)) but not of functional group richness ($F_{1,195} = 0.77$, $P = 0.380$). Examination of main effects shows that species richness effects are driven by site differences. Two and four species treatments were only carried out at the Swedish site and are responsible for the significant effects. Only the site–biomass interaction was significant ($F_{3,195} = 3.44$, $P < 0.018$ (Table 1)) and again reflects the significant site effects. Because of site differences, subsequent analyses of the effects of species richness, functional group richness and biomass are more usefully explored within individual sites (see Methods). The effects of diversity were not consistent between sites: at Gullmarsfjord (in Sweden) there were significant effects of species identity, species richness and functional group richness ($F_{9,77} = 7.12$, $P < 0.001$; $F_{1,77} = 5.31$, $P < 0.024$; and $F_{1,77} = 4.49$, $P < 0.037$, respectively (Table 1)). For the Ythan a and b experiments, there was no significant effect of species or functional group richness and only species identity was significant (Ythan a: $F_{3,19} = 4.4$, $P < 0.016$; Ythan b: $F_{3,40} = 50.92$, $P < 0.001$ (Table 1)), Boston bay was non-significant in all three components of diversity (Table 1). There were, however, consistent significant effects of biomass across all sites (Table 1). Biomass effects are not confounded by the excretion of metabolic nitrogen, which is trivial in our system.

As diversity increased, the coefficient of determination for each regression also increased, explaining a greater proportion of the

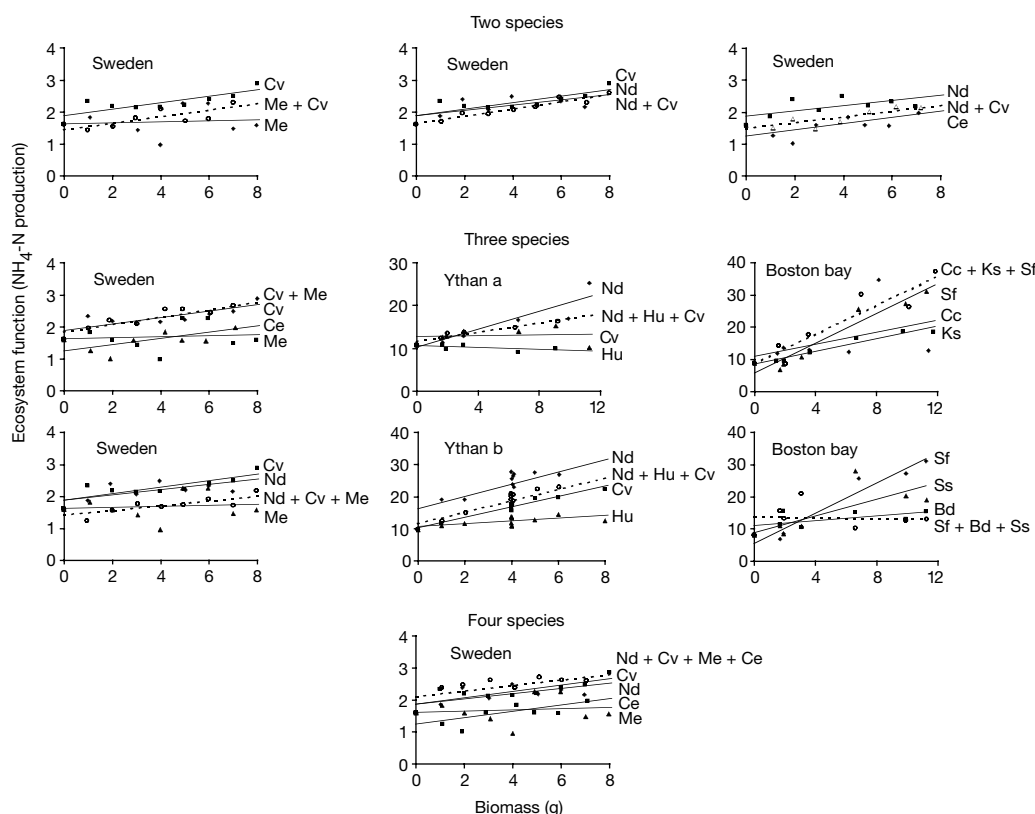


Figure 1 Biomass effects on infauna production for different species richness combinations. $\text{NH}_4\text{-N}$ production increases with biomass, there are variable effects of species richness on the magnitude or slope of regressions, and the responses are varied and different. Mixtures were plotted (dashed lines) against regressions obtained from

those single species contained in each mixture. Abbreviations: Cv, *Corophium volutator*; Nd, *Nereis diversicolor*; Hu, *Hydrobia ulvae*; Me, *Mytilus edulis*; Ce, *Cerastoderma edule*; Sf, *Salinator fragilis*; Ss, *Salinator solidus*; Bd, *Batillaria diemenensis*; Cc, *Callinassa ceramica*; and Ks, *Katylsia scalarina*. R^2 values for each regression are plotted in Fig. 2a.

variation in ecosystem function (Fig. 2a). For mixtures of two or three species, variance in the coefficient of determination for each regression was much less than for monocultures (Fig. 2b). This is consistent with the proposition that variability in function declines as species richness increases and hence ecosystem responses become more predictable^{1,6}.

Regressions obtained from single species treatments were used in an additive 'null model'²⁰ to generate expected regression slopes for comparison with observed multispecies treatments (see Methods). In only two instances were the observed mixture regressions significantly different from expected null regressions (significantly less for species *Salinator fragalis*, *Salinator solidus* and *Batillaria diemenensis*: $t = 3.45$, $P < 0.018$; significantly greater for species *Corophium volutator*, *Mytilus edulis* and *Cerastoderma edule*: $t = 2.73$, $P < 0.049$). The *S. fragalis*, *S. solidus* and *B. diemenensis* mixtures contained only one functional group (surficial deposit-feeding gastropods). The treatment was designed to be species rich but functionally poor. The *C. volutator*, *M. edulis* and *C. edule* mixture contained two functional groups (filter feeders *M. edulis* and *C. edule*, and a deposit feeder *C. volutator*). *C. volutator* resuspends a considerable amount of fine particulate material. The combination of resuspension and deposition, mediated by filtering activities of bivalves, may have contributed to the complementary cycling of material in this treatment. The lack of significant differences here is notable; diversity effects at these low levels of species richness seem to be additive rather than complementary. Less than 25% of all treatments showed evidence of over yielding (Fig. 3a), and thereby complementarity¹⁸. The percentage

of treatments showing over yielding increased with species richness (Fig. 3b). The high incidence of under yielding is also interesting and might be attributable to one of two factors: interference between species with similar niches or resource-use requirements²¹, or a reduction in the biomass or density of the best-performing species in mixture. There are other measures of over yielding^{3,18,19} but their calculation depends on the contribution of each species to the overall measure of function, which is not appropriate to flux measurements²⁰.

For the three sites (Udale bay, Culbin sands and Ythan estuary) where naturally occurring assemblages were used there was no significant difference in $\text{NH}_4\text{-N}$ production between sites ($F_{2,24} = 0.09$, $P = 0.911$). Sites were therefore treated as replicates for an analysis of covariance (ANCOVA) (both species richness and biomass covaried among treatments). There was no significant effect of biomass ($F_{1,24} = 3.85$, $P = 0.061$) but there was a highly significant effect of species richness ($F_{1,24} = 10.93$, $P < 0.003$ (Table 2)). Species richness varied among sites ($F_{2,24} = 3.81$, $P < 0.036$ (Table 2)) and all other interaction terms were non-significant. Backwards, stepwise multiple regression using individual species biomasses from each site as independent variables on

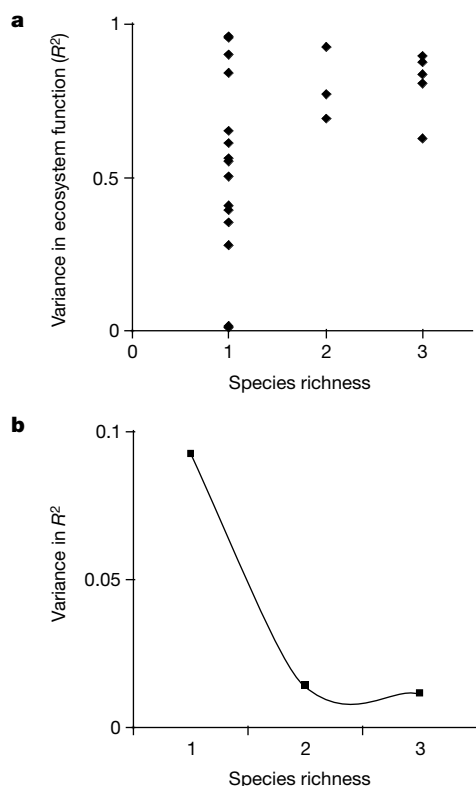


Figure 2 Patterns of variance. **a**, Coefficients of determination. R^2 values were obtained from each species-richness regression of $\text{NH}_4\text{-N}$ production on biomass. The R^2 values are plotted as a function of species richness. The amount of variability described by each regression increases as species richness increases, that is, between monocultures and species mixtures R^2 values generally increase. **b**, Variability in the coefficients of determination. As well as R^2 values increasing, the variability in the R^2 values (coefficient of variation) of mixtures declines relative to monocultures.

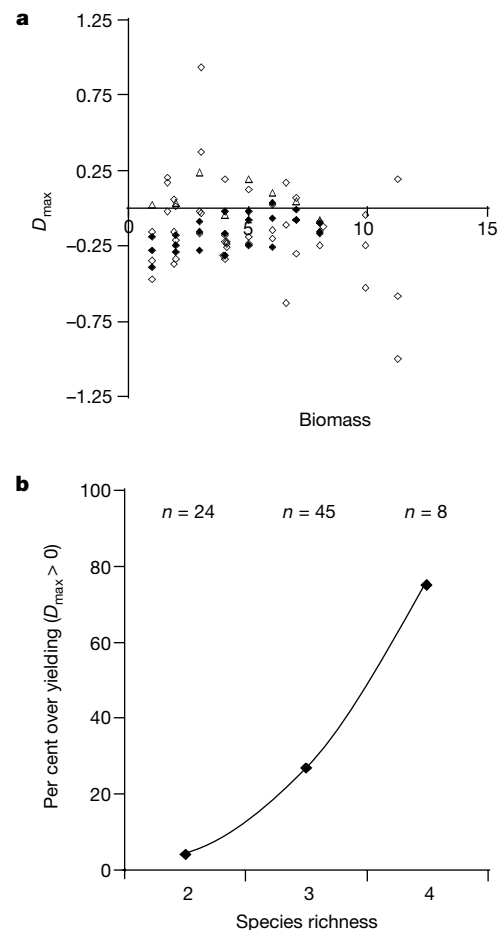


Figure 3 Diversity effects on over yielding. **a**, Over yielding (D_{max}) for all diversity treatments. Over yielding determines whether a species mixture outperforms the best single species treatment for those species contained in that mixture. For $\text{NH}_4\text{-N}$ production in the experiments detailed here less than 25% of all treatments over yield—most mixtures in fact under yield. **b**, Effects of species richness on over yielding. Over the gradient of species richness in this experiment (2–4 species), the percentage of treatments over yielding increases from 4 to 75%, respectively (although note the differences in sample size).

Table 2 Summary of ANCOVA and multiple regression analysis for three Scottish sites

ANCOVA	d.f.	s.s.	m.s.	F	P
Source of variation					
Site*	2	0.42	0.0003	0.09	0.911
Total Biomass†	1	0.16	0.016	3.85	0.061
Site*biomass	2	0.007	0.007	1.88	0.175
Species richness†	1	0.038	0.045	10.93	0.003
Site*species richness	2	0.012	0.016	3.81	0.036
Error	24	0.098	0.004		
Total	35	0.77			
Multiple regression analysis‡					
	Model	P	S.R.M.	R ²	Adj R ²
Udale Bay		0.001		0.89	0.87
<i>Nereis virens</i>	1.642	0.001	0.258	0.82	
<i>Hydrobia ulvae</i>	1.430	0.031	0.559	0.07	
Ythan Estuary		0.007		0.67	0.60
<i>Corophium volutator</i>	0.887	0.07	0.432	0.12	
<i>Nereis diversicolor</i>	17.48	0.004	4.48	0.55	
Culbin Sands					
<i>Macoma balthica</i>	1.3403	0.044	0.583	0.35	28.1

d.f., degrees of freedom. s.s., sum of squares. m.s., mean squares. Adj R², adjusted R².

* Site was treated as the main factor, with sites forming replicate assemblages.

† Both biomass and species richness covaried within and between different sites and were treated as covariates in this analysis.

‡ Multiple regression analysis for effects of individual species biomass on NH₄-N production.

NH₄-N production, demonstrated that at Udale *Nereis virens* had a significant effect on NH₄-N production ($P < 0.001$, $R^2 = 0.82$ (Table 2)). At Ythan *Nereis diversicolor* had disproportionate effects on NH₄-N production ($P < 0.004$, $R^2 = 0.55$ (Table 2)), whereas, at Culbin sands *Macoma balthica* was the only species to contribute significantly to variation in NH₄-N production ($P < 0.044$, $R^2 = 0.35$ (Table 2)). Thus, at each of the three sites individual species seem to be driving ecosystem function, that is, there was a sampling effect.

Our analysis shows that despite our extensive investigation we have been unable to detect a consistent effect of either species richness or functional group richness on ecosystem function among sites at a global scale. It is probable that these inconsistencies result from inherent site differences. The characteristic traits of those few species used in these artificially assembled communities will contribute to idiosyncratic patterns of ecosystem function with increasing diversity. Our large sample is due to replication of species richness treatments over a range of biomasses and it is under these conditions that we would expect to see idiosyncratic effects of diversity²². Our data are consistent with reduced variability (increased stability) in ecosystem function as diversity increases^{1,6} over the small range of species richness examined here. Also, there is evidence for a complementary effect of diversity on ecosystem function. Regionally there is much stronger evidence for effects of diversity from natural assemblages and our analysis shows how sampling effects can be revealed by appropriate experimental design. This last point is important, as the random generation of a diversity gradient has been criticized on the basis that natural communities do not randomly assemble²³, and that generating a diversity gradient in this way introduces the sampling effect. Here we show that species that contribute disproportionately to function are present across the range of species richness in natural assemblages and that increased function is due to the inclusion of less abundant species, whose contribution to function is more difficult to determine.

Although biodiversity affects the invasibility of marine systems²⁴, to our knowledge this is the first time that a diversity–function effect has been rigorously demonstrated for a marine benthic ecosystem. Such systems are under increasing anthropogenic pressures, including habitat loss and pollution. The availability of nutrients such as NH₄-N is of absolute importance for fuelling primary productivity within these systems, serving to maintain

their biodiversity and importance as nursery grounds for commercial fisheries species and migratory birds. Managing such systems in a way that will maintain their biodiversity minimizes the risk associated with predicting the variable effects of species loss. Idiosyncratic and consistent effects have previously been contrasted as opposing outcomes in biodiversity research. Our data show that they are not mutually exclusive and demonstrate a combination of variable effects arising from different species compositions, and a degree of consistency in biodiversity effects relating to complementarity and predictability. □

Methods

Azoic sediment and mesocosms

The sediment was frozen (−18 °C for two weeks), thawed and refrozen to kill infauna (animals living within the substrate) and most of the microbes. The defrosted sediment was then homogenized and distributed among the mesocosms. For Gullmarsfjord, sediment was also pre-sieved (1-mm mesh) before freezing.

Mesocosms consisted of 4-l aquaria containing 1.5 l of azoic sediment and 2 l of seawater (prefiltered through a 64-µm mesh). Each mesocosm was individually aerated, simultaneously mixing the water column.

Sampling procedures

Water samples were taken at 3-d intervals (20 ml with replacement) and each experiment was run for 15 d. This period of time was sufficient to allow faunal habituation and the development of a microbial population within sediments. Water samples were assayed on an FIA star 5010 flow analyser using standard operating procedures.

Species pools

Species used at the Ythan site included the polychaete *N. diversicolor*, the amphipod *C. volutator* and the gastropod *Hydrobia ulvae*. At the Gullmarsfjord site the epifaunal bivalve *M. edulis* and infaunal bivalve *C. edule* were used in treatments. The Boston bay site included the gastropods *S. fragilis*, *S. solidus* and *B. diemenensis*, which formed a three-species functional group of errant surficial deposit feeders, as well as the decapod *Callinassa ceramica* and the infaunal bivalve *Katylsia scalarina*. The polychaete *N. virens* and the deposit-feeding bivalve *M. balthica* were found to be dominant members of the fauna from natural assemblages.

Experimental design and analysis

We carried out two separate experiments: first, a direct manipulation of species richness in artificially assembled communities; and second, natural undisturbed communities were sampled for a greater range of species richness and to obtain small, delicate species that could not be manipulated directly. The dependent variable NH₄-N was log transformed to normalize the data. For the artificially assembled communities two analyses were carried out, the first design treated site as a factor; biomass, species richness and functional group richness were treated as covariates. Each site was then analysed separately (site effects were shown to be significant in the first analysis and were responsible for determining significant species richness and biomass terms). For the second design species identity (composition) was the main factor, whereas biomass, species richness and functional richness were treated as covariates for the purposes of regression analysis. For the naturally occurring communities, different sites formed replicate communities. Site was treated as the primary factor, whereas both biomass and species richness were treated as covariates.

Artificially assembled communities

Logistical constraints limited the number of assembled species richness treatments at different sites. At the Ythan site three-species treatments contained *N. diversicolor*, *H. ulvae* and *C. volutator* individually, and then in combination. Boston bay one-species treatments contained *S. fragilis*, *S. solidus*, *B. diemenensis*, *C. ceramica* and *K. scalarina* individually. The three-species treatments consisted of *S. fragilis*, *S. solidus*, *B. diemenensis* and *S. fragilis*, *C. ceramica*, *K. scalarina*, respectively. The Gullmarsfjord site featured 1–4 species treatments (*N. diversicolor*, *C. volutator*, *M. edulis* and *C. edule*, individually; (*N. diversicolor* and *C. edule*), (*N. diversicolor* and *C. volutator*) and (*C. volutator* and *M. edulis*) paired; (*C. volutator*, *M. edulis*, *C. edule*) and (*C. volutator*, *N. diversicolor*, *M. edulis*) in triads; with four species in the final treatment). Each species richness treatment had a nested set of biomass treatments that were fixed across each of these species richness levels. For mixtures, per cent composition was maintained while biomass increased. The rationale for this approach is that as the biomass or density of individual and component species in mixture increase so too will function.

Naturally assembled communities

Because of the small size and delicate nature of many invertebrates the *a priori* estimation of species biomass within ambient sediments is difficult. At each of the Scottish sites ambient communities were sampled and a biomass gradient established by sampling a range of increasing sediment surface areas, resulting in covariation between species richness and biomass through a sampling area effect. Animals were sieved from sediments

and placed in mesocosms. On termination of the experiment these macrofauna were retrieved and enumerated. We analysed the data by ANCOVA. The contribution of each species to variation in $\text{NH}_4\text{-N}$ production as biomass increased was assessed by multiple regression. This allows for the identification of species that might be regarded as causing a sampling effect, but does not allow for the determination of a diversity effect such as over yielding.

Null model

We used observations derived from single-species regression slopes to generate an additive null mixture regression slope against which to compare observed responses from multi species mixtures. The null model was of the form $E(j) = (\sum_{i=1}^n (\alpha_i p_i + b_i) - M_{\text{ZERO}}) + M_{\text{ZERO}}$ where $E(j)$ is the expected multispecies response at biomass j ; p_i is the proportion of species i represented in mixture by biomass; α_i is the slope of the regression line defining the relationship between biomass and ecosystem response for species i ; and b_i is the intercept of the regression line defining the relationship between biomass and ecosystem response for species i . M_{ZERO} is the mean, observed ecosystem response at zero biomass. It is important to distinguish between additive and substitutive experimental designs²⁵; the null model approach detailed here is of an additive statistical design, whereas calculation of metrics such as D_{Max} (over yielding) requires a substitutive experimental design. The experimental approach we have developed here allows for exploration of both approaches.

Received 24 November 2000; accepted 29 January 2001.

1. Naeem, S. & Li, S. Biodiversity enhances ecosystem reliability. *Nature* **390**, 507–509 (1997).
2. Tilman, D., Lehman, C. L. & Thompson, K. T. Plant diversity and ecosystem productivity: theoretical considerations. *Proc. Natl Acad. Sci. USA* **94**, 1857–1861 (1997).
3. Hooper, D. U. The role of complementarity and competition in ecosystem responses to variation in plant diversity. *Ecology* **79**, 704–719 (1998).
4. Hector, A. *et al.* Plant diversity and productivity experiments in European grasslands. *Science* **286**, 1123–1127 (1999).
5. Wardle, D. A., Bonner, K. I. & Barker, G. M. Stability of ecosystem properties in response to above-ground functional group richness and composition. *Oikos* **89**, 11–23 (2000).
6. McGrady-Steed, J., Harris, P. M. & Morrin, P. J. Biodiversity regulates ecosystem predictability. *Nature* **390**, 162–165 (1997).
7. McGrady-Steed, J. & Morin, P. J. Biodiversity, density compensation and the dynamics of populations and functional groups. *Ecology* **81**, 361–373 (2000).
8. Huston, M. *et al.* No consistent effect of plant diversity on productivity. *Science* **289**, 1255–1257 (2000).
9. Hector, A. *et al.* No consistent effect of plant diversity on productivity–response. *Science* **289**, 1257–1259 (2000).
10. Wardle, D. *et al.* Biodiversity and ecosystem function: an issue in ecology. *Bull. Ecol. Soc. Am.* **81**, 235–239 (2000).
11. Naeem, S. Reply to Wardle *et al.* *Bull. Ecol. Soc. Am.* **81**, 241–246 (2000).
12. Wardle, D. A., Bonner, K. I. & Nicholson, K. S. Biodiversity and plant litter: experimental evidence which does not support the view that enhanced species richness improves ecosystem function. *Oikos* **80**, 470–480 (1997).
13. Nixon, S. W. in *Estuaries and nutrients* (eds Neilson, B. J. & Cronin, L. E.) 111–138 (Humana, Clifton, 1981).
14. Nedwell, D. B., Jickells, T. D., Trimmer, M. & Sanders, R. Nutrients in estuaries. *Adv. Ecol. Res.* **29**, 43–92 (1999).
15. Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 1463–1468 (1999).
16. Ives, A. R., Gross, K. & Klug, J. L. Stability and variability in competitive communities. *Science* **286**, 542–544 (1999).
17. Hughes, J. B. & Roughgarden, J. Species diversity and biomass stability. *Am. Nat.* **155**, 618–627 (2000).
18. Loreau, M. Separating sampling and other effects in biodiversity experiments. *Oikos* **82**, 600–602 (1998).
19. Hector, A. The effect of diversity on productivity: detecting the role of species complementarity. *Oikos* **82**, 597–599 (1998).
20. Emmerson, M. C. & Raffaelli, D. G. Detecting the effects of diversity on measures of ecosystem function: experimental design, null models and empirical observations. *Oikos* **91**, 195–203 (2000).
21. Tilman, D. Causes, consequences and ethics of biodiversity. *Nature* **405**, 208–211 (2000).
22. Lawton, J. H., Naeem, S., Thompson, L. J., Hector, A. & Crawley, M. J. Biodiversity and ecosystem function: getting the Ecotron experiment in its correct context. *J. Func. Ecol.* **12**, 848–852 (1998).
23. Wardle, D. A. Is “sampling effect” a problem for experiments investigating biodiversity–ecosystem function relationships? *Oikos* **87**, 403–407 (1999).
24. Stachowicz, J. J., Whitlatch, R. B. & Osman, R. W. Species diversity and invasion resistance in a marine ecosystem. *Science* **286**, 1577–1579 (1999).
25. Jolliffe, P. A. The replacement series. *J. Ecol.* **88**, 371–385 (2000).

Supplementary information is available on Nature’s World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank L. Pihl and associated research staff at KMRS, Sweden and S. Hall and Flinders University for support in Australia. We also thank C. Biles and S. Way for help with fieldwork. Our work was supported by NERC and ESF grants to D.R. and D.P., and Carnegie trust to M.E.

Correspondence and requests for materials should be addressed to M.C.E. (e-mail: m.emmerson@abdn.ac.uk).

Larval dispersal potential of the tubeworm *Riftia pachyptila* at deep-sea hydrothermal vents

Adam G. Marsh^{*†}, Lauren S. Mullineaux[‡], Craig M. Young[§] & Donal T. Manahan^{*}

^{*} Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA

[‡] Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

[§] Division of Marine Science, Harbor Branch Oceanographic Institution, Fort Pierce, Florida 34946, USA

Hydrothermal vents are ephemeral because of frequent volcanic and tectonic activities associated with crust formation^{1–3}. Although the larvae of hydrothermal vent fauna can rapidly colonize new vent sites separated by tens to hundreds of kilometres^{4,5}, the mechanisms by which these larvae disperse and recruit are not understood. Here we integrate physiological, developmental and hydrodynamic data to estimate the dispersal potential of larvae of the giant tubeworm *Riftia pachyptila*. At *in situ* temperatures and pressures (2 °C and 250 atm), we estimate that the metabolic lifespan for a larva of *R. pachyptila* averages 38 days. In the measured flow regime at a fast-spreading ridge axis (9° 50' N; East Pacific Rise), this lifespan results in potential along-ridge dispersal distances that rarely exceed 100 km. This limited dispersal results not from the physiological performance of the embryos and larvae, but instead from transport limitations imposed by periodic reversals in along-ridge flows and sustained episodes of across-ridge flow. The lifespan presented for these larvae can now be used to predict dispersal under current regimes at other hydrothermal vent sites.

Larval dispersal remains one of the most perplexing problems for understanding how hydrothermal vent organisms colonize vent sites^{4,6}. We have surmounted a significant challenge to addressing this problem by successfully rearing embryos of *Riftia pachyptila* (Polychaeta: Siboglinidae) to the larval stage under deep-ocean temperatures and pressures. Using custom designed culture systems (Teflon-coated, stainless steel chambers with a continuous flow of seawater at 2 °C and 250 atm) we have followed the early development of *R. pachyptila* embryos and larvae.

Early cleavage (Fig. 1a–d) was similar to that of bathyal (600 m) siboglinids⁷, except that embryos of *R. pachyptila* require higher pressure. After 34 d of development, we observed a swimming trochophore larval stage with two equatorial bands of simple cilia (Fig. 1e). As in the non-feeding larvae of another vestimentiferan worm, *Lamellibranchia* sp.⁷, larvae of *R. pachyptila* did not develop a mouth through the first 34 d of development. Once cleavage was initiated (after the exclusion of two polar bodies), cells divided at an average rate of 1.8 d per division (Fig. 2a). Some fertilized eggs were returned to the vent site at 2,500 m in mesh-covered 50-ml culture tubes to develop *in situ* on the sea floor (2 °C). Rates of early development *in situ* were equivalent to those observed in pressure chambers (Fig. 2a).

Larvae of Antarctic echinoderms show a remarkable metabolic efficiency with larval lifespans potentially exceeding 1 yr at low temperatures (–1.5 °C)^{8,9}. Larvae of hydrothermal vent organisms dispersing between vents at 2 °C might also have long metabolic lifespans. Larval lifespan is determined in part by the rate of energy use during development and the initial amount of energy available

[†] Present address: College of Marine Studies, University of Delaware, Lewes, Delaware 19958, USA.

in the egg. We measured metabolic rates at 2 °C and 250 atm in early embryos of *R. pachyptila* by end-point respirometry¹⁰. Developmental stages of *R. pachyptila* respired 5.1 ± 0.3 (mean $\pm$ s.e.) pmol O₂ per h per embryo and later stages (32–64 cells) averaged 17.8 ± 1.4 pmol O₂ per h per embryo (Fig. 2b). Biochemical analyses of the protein and lipid-class composition of eggs of *R. pachyptila* were conducted on samples from several females collected at different times of the year (Fig. 3a). To estimate a metabolic lifespan for developmental stages, we assumed that 50% of an egg's protein mass (34.5 ± 7.8 ng; mean $\pm$ s.d.) and 90% of its wax-ester lipid mass (144.4 ± 19.1 ng) could be metabolized (wax esters are used exclusively for energy storage in cold-water invertebrates)^{11–13}. The average metabolic energy available for development was calculated to be 5.5 ± 0.8 μ J per egg (mean $\pm$ s.d., range 4.7–7.1 μ J). Balancing the cumulative rate of energy use during early development (Fig. 2b) with available energy, a typical larva of *R. pachyptila* could potentially survive for 38 d (range 34–44 d).

Knowing the potential metabolic lifespan of a larva of *R. pachyptila* solves only part of the dispersal distance puzzle. The solution also depends on the vertical and horizontal directions and distances larvae travel. Passive vertical movement of the eggs and early embryos of *R. pachyptila* has been suggested to be high

(28 m d^{-1} , measured at 102 atm)¹⁴. However, our direct measurements of buoyancy in isolated pressure vessels with Plexiglass view-ports revealed that eggs are near-neutrally buoyant at 250 atm (and 2 °C), with an average vertical velocity of only 2 m d^{-1} (Fig. 3b). At active hydrothermal venting sites, hot effluents mix with seawater to form buoyant plumes that rise until they cool to neutral buoyancy at a distance of 175–200 m above the bottom (in less than a day)^{15,16}. Many early life stages of hydrothermal vent animals have been found entrained in this neutrally buoyant plume^{17,18}, and it is likely that eggs of *R. pachyptila* become entrained in this water mass as well.

At the height of the buoyant plume, the flow regime at our site (9° 50' N, East Pacific Rise; EPR) was assessed over a period of one year with two long-term current records (April to December 1999 and December 1999 to April 2000) from an Aanderaa current meter positioned 175 m above bottom on the ridge axis near the 'Biovent' site (9° 50.9' N, 104° 17.6' W). The current velocities were variable, with periodicities at semi-diurnal, diurnal and fortnightly time-scales. Residual flows tended to be oriented roughly along the

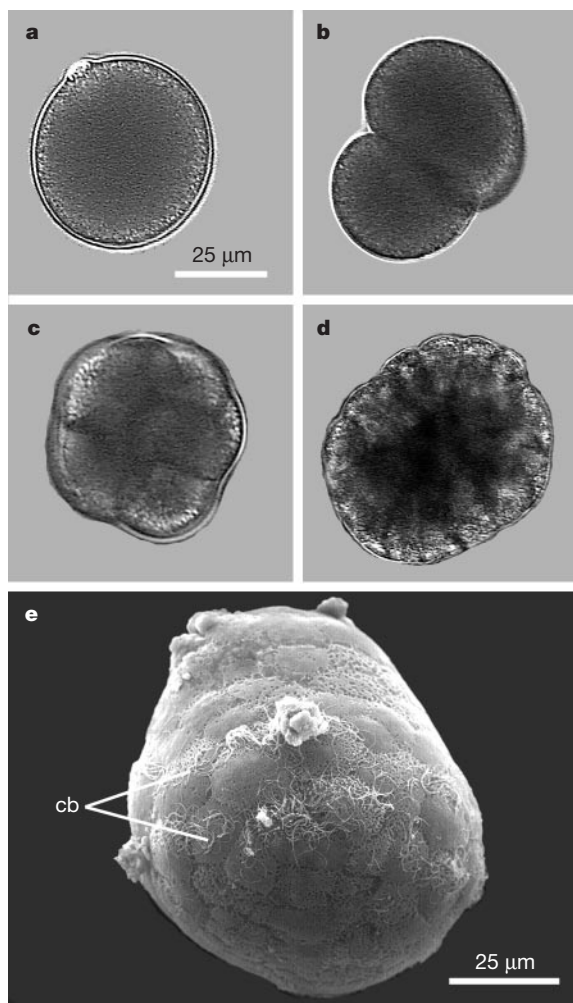


Figure 1 Early development to a larval stage in *Riftia pachyptila* cultured in pressure vessels at 2 °C and 250 atm. **a–d**, Light micrographs of first polar body extrusion (1 day; **a**), unequal blastomeres at the two-cell stage after polar lobe formation and first cleavage (3 days; **b**), the eight-cell stage (7 days; **c**), and a 128+ cell embryo (14 days; **d**). **e**, A scanning electron micrograph of a 34-day larva reveals two equatorial ciliary bands (cb) but no evidence of a mouth, apical tuft or other distinct morphological features.

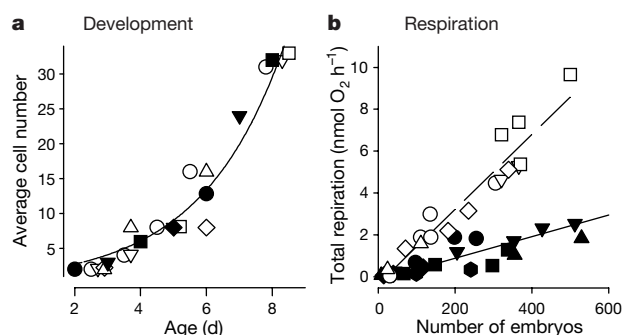


Figure 2 The developmental and respiration rates of *Riftia pachyptila* embryos. Symbols represent different cultures (different females, different cruises). **a**, For development, no difference was evident in cell division rates between cultures in pressure vessels (open symbols) or the *in situ* culture tubes (filled symbols). The combined regression of cell number against time ($y = 1.0283e^{(0.4299x)}$; $r^2 = 0.9623$) reveals a cleavage rate of 1.8 d per division. **b**, Embryo respiration measurements are divided into early (4–16 cells, filled symbols) and late (32–64 cells, open symbols) stages. Respiration rates of an individual embryo are determined as the slope of the linear regression equations: early $R = -0.147 + 0.00514x$, $r^2 = 0.9580$, $n = 19$; late $R = -0.354 + 0.0179x$, $r^2 = 0.9138$, $n = 18$. For later stages (> 20 d) a maximum metabolic rate of 21 pmol O₂ per h per individual was estimated from the rate of change in respiration during earlier development.

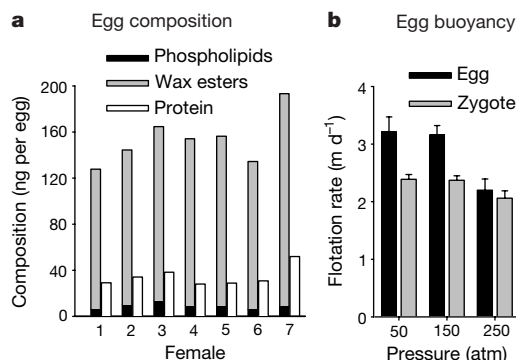


Figure 3 Biochemical composition and buoyancy of the eggs of *Riftia pachyptila*. **a**, Egg composition was assessed in different females, one from a cruise in May 1998, four from November 1998, and two from December 1999. Standard errors were less than 6% for the lipid assays and less than 2% for the protein assays, but are not plotted on the stacked histograms. **b**, Buoyancy was measured as flotation rates in pressurized vessels equipped with view ports for monitoring the vertical position of the eggs and zygotes (mean $\pm$ s.e.; sample sizes ranged from 62 to 223 observations per treatment).

NNW/SSE trending axis. The four-month current record from December 1999 to April 2000 gave evidence of a longer potential dispersal distance for larvae of *R. pachyptila*, which we have selected to present in detail.

To estimate dispersal potential of *R. pachyptila*, we modelled the movement of neutrally buoyant particles ('larvae') released at hourly intervals into the observed current regime. We followed

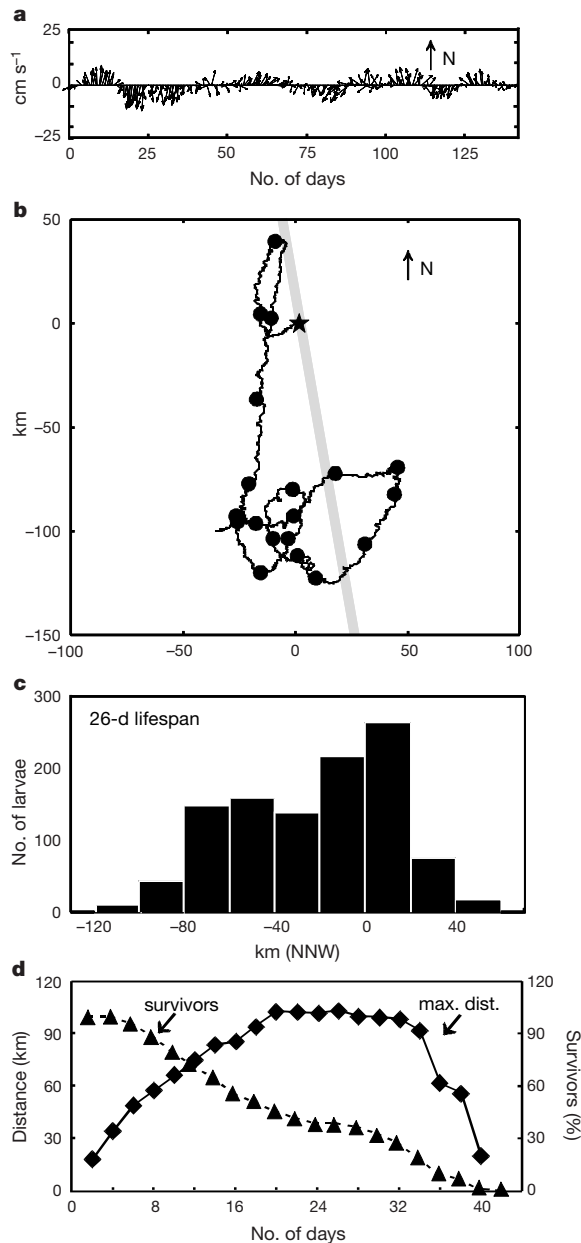


Figure 4 Dispersal potential of larvae of *Riftia pachyptila* modelled from current regimes at 9° N East Pacific Rise. **a**, Current velocities represented by arrows showing compass heading and velocity. **b**, Plan view of the trajectory of a water parcel, with its origin (star) at the location of the current meter and its path calculated from hourly-averaged velocities (filled circles denote positions at weekly intervals; grey line denotes the ridge axis). **c**, Along-axis distances (NNW is positive) travelled by larvae with a selected (26-d) lifespan released at hourly intervals in the 9° N current regime. Out of the 3,388 larval releases modelled, only 39% remained within 25 km of the ridge axis (these were termed 'survivors'). The maximum dispersal distance for a survivor with this 26-d lifespan was 103 km SSE. **d**, Maximum distance dispersed and percentage of survivors for larvae of each specified lifespan released in the dispersal model as described in **c**. For a lifespan greater than 42 d, no larval releases remained within 25 km of the ridge crest.

each larval trajectory (Fig. 4b) for either a specified lifespan or until a larva had moved more than 25 km off axis (at which point it was considered 'lost' from the ridge system). During the four-month interval from December 1999 to April 2000 (Fig. 4a, b), larvae released at most time points experienced substantial cross-axis transport, with the exception of those released between 16 and 42 d after the current meter was used, when flow was consistently aligned SSW, roughly along the axis (Fig. 4b). Only a few combinations of release time and lifespan resulted in net dispersal distances greater than 100 km (only 0.9% of the releases); these occurred for larvae released on or shortly after day 16 with lifespans of 20–34 d (Fig. 4d). The maximum along-axis dispersal distance for any individual was 103.4 km SSE, and the furthest NNW any individual travelled was 47.3 km. In the eight-month current profile from April 1999 to December 1999, the longest dispersal distance SSE along-axis was only 54 km. Notably, the along-ridge flow reversals at our 9° N site result in a much higher probability that larvae will be retained within a few tens of kilometres of their source population, which might provide a mechanism of larval retention and rapid population growth at this site⁵.

Contrary to expectations, a longer metabolic lifespan at 9° N EPR did not necessarily translate into a longer along-ridge dispersal distance. The more time a larva spent in the currents, the more likely it was to be swept off axis (Fig. 4d, 'survivors' curve). Sustained intervals of off-axis flow are characteristic of other sites on Eastern Pacific ridges^{19,20}, indicating that long-lived larvae probably have similar fates at these sites. Furthermore, mean along-axis velocities for extended (≥ 40 d) intervals are typically less than 3 cm s⁻¹ on Eastern Pacific ridges because of reversals in current direction^{19–22}. These velocities would result in dispersal distances of less than 100 km for any larvae of *R. pachyptila* that remained near the ridge. An exception to this hydrodynamic limitation of dispersal occurs at 13° N EPR where periods of unidirectional, along-axis flow with mean velocity of 4.5 cm s⁻¹ can last as long as 60 d (ref. 22). In this region, a longer lifespan of *R. pachyptila* larvae would result in substantially increased (up to 245 km) dispersal distances. It is also important to note that while off-axis flows impede dispersal along a ridge, they may be responsible for rare but important dispersal events to isolated vents (for example, on seamounts) or between ridges. In these cases, larval lifespan would limit the range and influence the frequency of successful colonization. Our data show the overriding importance of current flow in determining dispersal potential and suggest that populations at different vent sites may have different dispersal limits depending on local current conditions.

We have integrated physiological, developmental and hydrodynamic data into a specific dispersal estimate for larvae of *R. pachyptila*. Our unique finding is a close relationship between the physiology of tubeworm larvae and the physics of the oscillatory current regimes at the 9° 50' N EPR site. There, the balance between larval lifespan and flow reversals can account for both larval dispersal to colonize new vent sites, and larval retention to rapidly populate an existing vent site. At our site in this region it does not appear that the dispersal distance of *R. pachyptila* is limited by the physiological performance of its larvae, but by temporal oscillations in the along-axis currents and by larval loss in cross-axis flows. Now that we have provided an estimate of the metabolic lifespan for larvae of *R. pachyptila*, dispersal distances can be predicted under the current regimes at other hydrothermal vent sites where this species occurs. □

Methods

Culture methods and measurements

Adult *Riftia pachyptila* were collected from vent sites along the EPR (9° 50' N) and returned to the surface. In 2–3 h of arrival aboard ship, ripe eggs and sperm were removed from the gonoducts of adults at ambient pressure (1 atm), and combined for fertilization at 2 °C. The highest fertilization success rates (> 90%) were obtained by keeping

concentrated eggs and sperm together ($\sim 1,000 \text{ ml}^{-1}$) for 2–3 h, before transferring them to the pressure chambers at $50\text{--}100 \text{ ml}^{-1}$. We used standard high-performance liquid chromatography pumps and inert fittings to establish flow-through cultures in which filtered seawater was continuously replaced in a pressure vessel¹³. At the flow rates used, 90% of the column volume was exchanged in a 36-h period (that is, 1 ml min^{-1} for a 500-ml vessel with internal mixing, as verified by direct calibration). Embryos and larvae could be reared for up to 34 d using this culture system. We measured respiration rates of developing embryos using end-point assays. Embryos were transferred and maintained at pressure (250 atm) in 1-ml pressure vessels during 4–6 h incubation periods. The vessels were then depressurized and the oxygen concentration ($p\text{O}_2$) in the water measured at 1 atm using a microcathode oxygen sensor^{9,10}. Under our conditions there was no degassing when the sample chambers were depressurized because the initial seawater oxygen concentration (O_2 solubility at 1 atm) does not increase when pressurized in our sealed system. Respiration rates were determined by transferring under pressure different numbers of embryos into the 1-ml pressure vessels, and then regressing the resultant total oxygen use against total number of individuals (Fig. 2b). The linear response over a broad range of embryo concentrations verifies that respiration rates were not impacted by concentration-dependent interactions. This method of analysis also controls for background respiration rates in the 1-ml chambers by using only the regression slope¹⁰. In addition, we have made respiration measurements using the low oxygen concentrations of deep-bottom water at vent sites ($170 \mu\text{M O}_2$ measured in samples collected by *Alvin* in Niskin bottles) and have found no effects on larval metabolic rates. We measured the protein content of eggs by Bradford assay with bovine serum albumin as the standard. Lipid class analyses were performed using thin layer chromatography and quantification by flame-ionization detection (Iatroscan)²⁴.

Buoyancy of eggs and zygotes was measured in pressurized 80-ml volume stainless steel vessels with an inside diameter of 3.2 cm. Plexiglas ports on both the top and bottom sides permitted illumination and visualization of the central 1-cm diameter portion of the water column using a dissecting microscope. All observations were done under pressure at 2°C . After all eggs had floated to one end of the chamber, we reversed the chamber rapidly and took samples every 5 min over the following hour. At each sampling interval, the number of eggs arriving at the upper viewport was counted. Edge effects were avoided by counting only those eggs visible in the middle third of the column.

Dispersal model

The current velocities (corrected for magnetic variation) recorded at our study site varied strongly on fortnightly periods and this data record was sufficiently long to realize eight full cycles of along-axis current reversals (December 1999 to April 2000). The neutrally buoyant plume at this site was centred roughly 175-m off bottom (as reported previously¹⁵ and as verified by our own measurements of conductivity, temperature and density (CTD/transmissometer casts). Positioning the current meters at this depth provided measurements of current regimes that were the most relevant to larval dispersal. The ridge trends NNW/SSE, at a 10° angle from geographic North, requiring the currents used for the dispersal calculations to be translated into along- and across-axis alignments. In the dispersal model, the along-axis distance x_i (in m) travelled by an individual larva released at time step i was calculated as

$$x_i = dt \sum_{j=i}^{i+l-1} u_j \quad \text{for } i = 1 \text{ to } (n-l+1)$$

where dt is a 1-h time step, u_j is the mean along-axis velocity (m h^{-1}) during j th 1-h time step, n is the number of 1-h intervals in the current record, and l is the larval lifespan (h). The cross-axis distance y_i was calculated similarly, substituting the cross-axis velocity v_j for u_j . We set a limit of 25 km as the maximum distance that a larva could travel off axis and still be able to colonize the ridge. This distance corresponds roughly to the outer limit of the ridge base, and was selected because the ridge-affected flows tend to extend roughly to this distance^{19,25,26}. Any larva that strayed more than 25 km off axis was excluded from further analysis.

Received 21 November 2000; accepted 1 February 2001.

1. Fornari, D. J. *et al.* Time-series temperature measurements at high-temperature hydrothermal vents, East Pacific Rise $9^\circ 49\text{--}51^\circ \text{N}$: monitoring a crustal cracking event. *Earth Planet. Sci. Lett.* **160**, 419–431 (1998).
2. Fornari, D. J. & Embley, R. W. In *Physical, Chemical, Biological and Geological Interactions within Seafloor Hydrothermal Systems* (eds Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. & Thomson, R. E.) 1–46 (American Geophysical Union Monograph, Washington DC, 1995).
3. Haymon, R. M. *et al.* Volcanic eruption of the mid-ocean ridge along the East Pacific Rise crest at $9^\circ 45\text{--}52^\circ \text{N}$: direct submersible observations of sea-floor phenomena associated with an eruption event in April, 1991. *Earth Planet. Sci. Lett.* **119**, 85–101 (1993).
4. Mullineaux, L. S., Fisher, C. R., Peterson, C. H. & Schaeffer, S. W. Tubeworm succession at hydrothermal vents: use of biogenic cues to reduce habitat selection error? *Oecologia* **123**, 275–284 (2000).
5. Lutz, R. A. *et al.* Rapid growth at deep-sea vents. *Nature* **371**, 663–664 (1994).
6. Tyler, P. A. & Young, C. M. Reproduction and dispersal at vents and cold seeps. *J. Mar. Biol.* **79**, 193–208 (1999).
7. Young, C. M., Vazquez, E., Metaxas, A. & Tyler, P. A. Embryology of vestimentiferan tube worms from deep-sea methane/sulphide seeps. *Nature* **381**, 514–516 (1996).
8. Shilling, F. M. & Manahan, D. T. Energy metabolism and amino acid transport during early development of Antarctic and temperate echinoderms. *Biol. Bull.* **187**, 398–407 (1994).
9. Marsh, A. G., Leong, P. K. K. & Manahan, D. T. Energy metabolism during embryonic development and larval growth of an Antarctic sea urchin. *J. Exp. Biol.* **202**, 2041–2050 (1999).
10. Marsh, A. G. & Manahan, D. T. A method for accurate measurements of the respiration rates of marine invertebrate embryos and larvae. *Mar. Ecol. Prog. Ser.* **184**, 1–10 (1999).

11. Evanson, M., Bornhold, E. A., Goldblatt, R. H., Harrison, P. J. & Lewis, A. G. Temporal variation in body composition and lipid storage of the overwintering, subarctic copepod *Neocalanus plumchrus* in the Strait of Georgia, British Columbia (Canada). *Mar. Ecol. Prog. Ser.* **192**, 239–247 (2000).
12. Hagen, W. & Kattner, G. Lipid metabolism of the Antarctic euphausiid *Thysanoessa macrura* and its ecological implications. *Limnol. Oceanogr.* **43**, 1894–1901 (1998).
13. Kattner, G. & Hagen, W. Lipid metabolism of the Antarctic euphausiid *Euphausia crystallorophias* and its ecological implications. *Mar. Ecol. Prog. Ser.* **170**, 203–213 (1998).
14. Cary, S. C., Felbeck, H. & Holland, N. D. Observations on the reproductive biology of the hydrothermal vent tube worm *Riftia pachyptila*. *Mar. Ecol. Prog. Ser.* **52**, 89–94 (1989).
15. Baker, E. T. *et al.* Hydrothermal plumes along the East Pacific Rise, $8^\circ 40'$ to $11^\circ 50' \text{N}$: 1. Plume distribution and relationship to the apparent magmatic budget. *Earth Planet. Sci. Lett.* **128**, 1–17 (1994).
16. Lupton, J. E. *et al.* Tracking the evolution of a hydrothermal event plume with a RAFOS neutrally buoyant drifter. *Science* **280**, 1052–1055 (1998).
17. Kim, S. L., Mullineaux, L. S. & Helfrich, K. Larval dispersal via entrainment into hydrothermal vent plumes. *J. Geophys. Res.* **99**, 12655–12665 (1994).
18. Mullineaux, L. S., Wiebe, P. H. & Baker, E. T. Larvae of benthic invertebrates in hydrothermal vent plumes over Juan de Fuca Ridge. *Mar. Biol.* **122**, 585–596 (1995).
19. Thomson, R. E., Roth, S. E. & Dymond, J. Near-inertial motions over a mid-ocean ridge: effects of topography and hydrothermal plumes. *J. Geophys. Res.* **95**, 12961–12966 (1990).
20. Cannon, G. A. & Pashinski, D. J. Variations in mean currents affecting hydrothermal plumes on the Juan de Fuca Ridge. *J. Geophys. Res.* **102**, 24965–24976 (1997).
21. Crane, K., Aikman III, F. & Foucher, J.-P. The distribution of geothermal fields along the East Pacific Rise from $13^\circ 10' \text{N}$ to $8^\circ 20' \text{N}$: implications for deep seated origins. *Mar. Geophys. Res.* **9**, 211–236 (1988).
22. Chevallon, P., Jollivet, D., Vangriesheim, A. & Desbruyères, D. Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 1. Influence of vent site distribution, bottom currents, and biological patterns. *Limnol. Oceanogr.* **42**, 67–80 (1997).
23. Jannasch, H. W., Wirsén, C. O. & Doherty, K. W. A pressurized chemostat for the study of marine barophilic and oligotrophic bacteria. *Appl. Environ. Microbiol.* **62**, 1593–1596 (1996).
24. Miller, C. B., Morgan, C. A., Pahl, F. G. & Sparrow, M. A. Storage lipids of the copepod *Calanus finmarchicus* from Georges Bank and the Gulf of Maine. *Limnol. Oceanogr.* **43**, 488–497 (1998).
25. Thomson, R. E., Gordon, R. L. & Dymond, J. Acoustic doppler current profiler observations of a mid-ocean ridge hydrothermal plume. *J. Geophys. Res.* **94**, 4709–4720 (1989).
26. Cannon, G. A., Pashinski, D. J. & Lemon, M. R. Mid-depth flow near hydrothermal venting sites on the southern Juan de Fuca ridge. *J. Geophys. Res.* **96**, 12815–12831 (1991).

Acknowledgements

We thank the captain and crew of the RV *Atlantis II*, and the support crew of the DSV *Alvin*. We acknowledge the data collection and assistance of S. Brooke, A. Green, S. Mills, M. Moore and D. Pace. C. Allen, S. Beaulieu, T. Griffin, H. Hunt, A. Metaxas, P. Tyler and J. Welch provided assistance at sea; M. Grosenbaugh provided programming and conceptual assistance with the dispersal model. This work was supported by grants from the National Science Foundation.

Correspondence and requests for materials should be addressed to D.T.M. (e-mail: manahan@usc.edu).

Foci of orientation plasticity in visual cortex

Valentin Dragot*, Casto Rivadulla*† & Mriganka Sur*

* Department of Brain and Cognitive Sciences and Center for Learning and Memory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Cortical areas are generally assumed to be uniform in their capacity for adaptive changes or plasticity^{1–4}. Here we demonstrate, however, that neurons in the cat striate cortex (V1) show pronounced adaptation-induced short-term plasticity of orientation tuning primarily at specific foci. V1 neurons are clustered according to their orientation preference in iso-orientation domains⁵ that converge at singularities or pinwheel centres^{6,7}. Although neurons in pinwheel centres have similar orientation tuning and responses to those in iso-orientation domains, we find that they differ markedly in their capacity for adaptive changes.

† Present address: Neuroscience and Motor Control Group (NEUROcom), E U Fisioterapia, Campus de Oza, 15006 A Coruña, Spain.

Adaptation with an oriented drifting grating stimulus alters responses of neurons located at and near pinwheel centres to a broad range of orientations, causing repulsive shifts in orientation preference and changes in response magnitude. In contrast, neurons located in iso-orientation domains show minimal changes in their tuning properties after adaptation. The anisotropy of adaptation-induced orientation plasticity is probably mediated by inhomogeneities in local intracortical interactions that are overlaid on the map of orientation preference in V1.

Cortical neuron responses are influenced significantly by the inputs that they receive from other neurons in their local network⁸. The structure of the orientation map in V1 implies that the orientation distribution of local connections would vary with a neuron's position within the map: neurons in pinwheel centres are likely to be connected to neurons of a broader range of orientations than neurons in iso-orientation domains. Despite the difference in local inputs to neurons in pinwheel centres and iso-orientation domains, both these classes of cell have been described to have similar orientation tuning characteristics⁹. However, orientation tuning, as such, may not fully reflect the difference between the properties of neurons across V1. Indeed, if orientation selectivity relies significantly on the alignment of thalamic afferents^{10–12} (see also ref. 13), a high degree of precision of thalamocortical projections to pinwheel centres or iso-orientation domains could account

for the sharpness of their orientation tuning, despite differences in the distribution of intracortical inputs in these regions. We reasoned that altering the efficacy of intracortical orientation-specific inputs to neurons in different locations of the orientation map might induce changes in the tuning properties of neurons in a manner that depends on cortical location.

We induced short-term plasticity in V1 neurons by adapting them to a stimulus of fixed orientation¹⁴. Adaptation affects a broad range of orientations by selectively reducing responses at stimuli at and near the adapting orientation^{15,16} and enhancing the response at other orientations¹⁴, thereby causing transient changes in orientation selectivity. We investigated the relationship between this form of orientation tuning plasticity and a neuron's location in the orientation preference map in V1 of adult cats. We used optical imaging of intrinsic signals to obtain the orientation map in a patch of V1 (Fig. 1a). We used the vascular pattern of the cortical surface in relation to the orientation map (Fig. 1b) to guide electrode penetrations aimed at iso-orientation domains or pinwheel centres. As pinwheel centres are locations where the preferred orientation of neurons changes rapidly, we determined an orientation gradient map as the two-dimensional spatial derivative at each pixel to identify these foci. The gradient map (Fig. 1c) shows that pinwheel centres are included in regions with the highest rate of orientation change, whereas the gradient is low in iso-orientation domains.

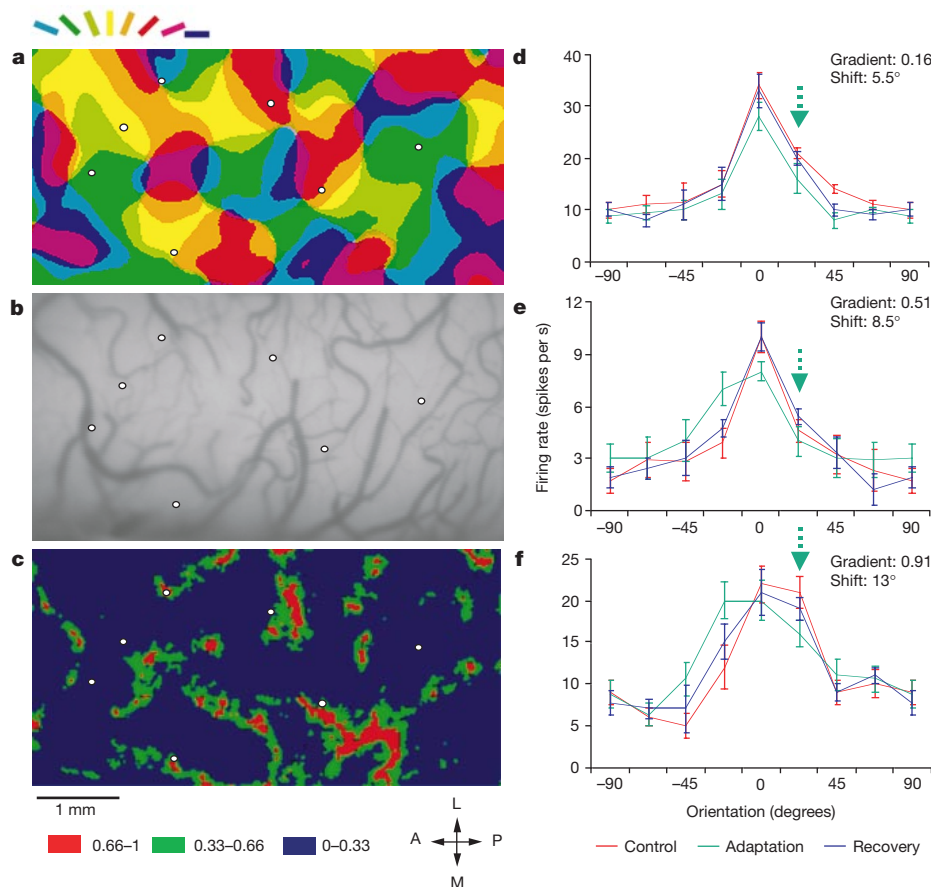


Figure 1 Adaptation-induced plasticity of orientation tuning and the orientation architecture of V1. **a**, Composite orientation map obtained by intrinsic signal imaging. The angle of preferred orientation of each pixel is shown according to the colour key (top). The map was smoothed using a low-pass filter (5×5 pixels). Circles indicate the location of seven representative neurons (out of 40 recorded) to illustrate the range of orientation and gradient distributions. **b**, Vascular pattern of the cortical surface for the region shown in **a**. **c**, Orientation gradient map in which gradient was distinguished into the following ranges:

red (0.66–1), green (0.33–0.66) and blue (<0.33). A, anterior; L, lateral; M, medial; R, posterior. **d–f**, Orientation tuning curves of three representative cells during control, adaptation and recovery conditions. The control optimal orientation is represented as 0° , and all subsequent tuning curves (during adaptation and recovery) are represented relative to the control condition. The adapting orientation is indicated by the green arrow. Each point in panels **d**, **e** and **f** represents mean value $\pm$ s.e.m.

These analyses yielded reliable maps of the orientation gradient, with the mean value of the gradient at each pixel stable to within 5.5% for maps obtained several hours apart ($n = 3$ animals, analysing pixels in each animal imaged 10 h apart; animal 1: 31,824 pixels spanning $2.08 \times 2.72 \text{ mm}^2$; animal 2: 75,888 pixels spanning $2.72 \times 4.96 \text{ mm}^2$; animal 3: 46,800 pixels spanning $2.08 \times 4 \text{ mm}^2$). The gradient map was used to relate plasticity in orientation preference of V1 neurons to the exact value of the local gradient at the recording site.

It has been shown^{14,17,18} that even brief periods of adaptation (lasting for a fraction of a second to a few seconds) can induce consistent changes in orientation selectivity and response magnitude of V1 neurons. For this study, we chose adaptation of 2-min duration to induce large shifts in orientation and response. Figure 1d–f demonstrates the relationship between location within the orientation map and adaptation-induced plasticity of orientation tuning for representative neurons. Adaptation to a given orientation induces a repulsive shift in orientation preference away from the adapting stimulus. Notably, the higher the value of the orientation gradient at the recording site, the larger is the magnitude of the shift in preferred orientation.

Figure 2 characterizes the plasticity of orientation tuning for a population of 118 V1 neurons (26, 52 and 40 cells in 3 animals). The proportion of cells showing significant shifts in preferred orienta-

tion increases as the orientation gradient increases. Figure 2a shows that 88% of the cells located at regions of the highest orientation gradient (>0.75) exhibit significant repulsive shifts after adaptation ($P < 0.05$, Student's t -test; on the basis of a trial-by-trial comparison between control and adaptation conditions). In contrast, 12% of the cells located in the lowest gradient areas show a significant shift in orientation preference. Figure 2b shows that the magnitude of the shift in optimal orientation is a monotonic function of the orientation gradient at each recording location (correlation coefficient $r = 0.6$, $P < 0.00005$, Pearson test; positive shifts are shown as repulsive with respect to the adapting orientation).

As neurons in pinwheel centres are located in close proximity to neurons of all orientations, we reasoned that, close to these centres, pronounced changes in orientation tuning could be induced after adaptation to a broad range of orientations. We studied cells in all gradient regions with a wide range of orientation differences between the cell's optimal orientation and that of the adapting stimulus ($\Delta\theta$)—the $\Delta\theta$ values were statistically independent of the orientation gradient values ($r = 0.049$, $P > 0.1$). Cells located at high gradient regions exhibited significant shifts in preferred orientation for a broader range of $\Delta\theta$ compared with cells in low-gradient domains (Fig. 2c). The magnitude of the shift in preferred orientation is smaller as $\Delta\theta$ increases¹⁴; here we show that for large $\Delta\theta$ ($>45^\circ$), the orientation shift is smaller in all the

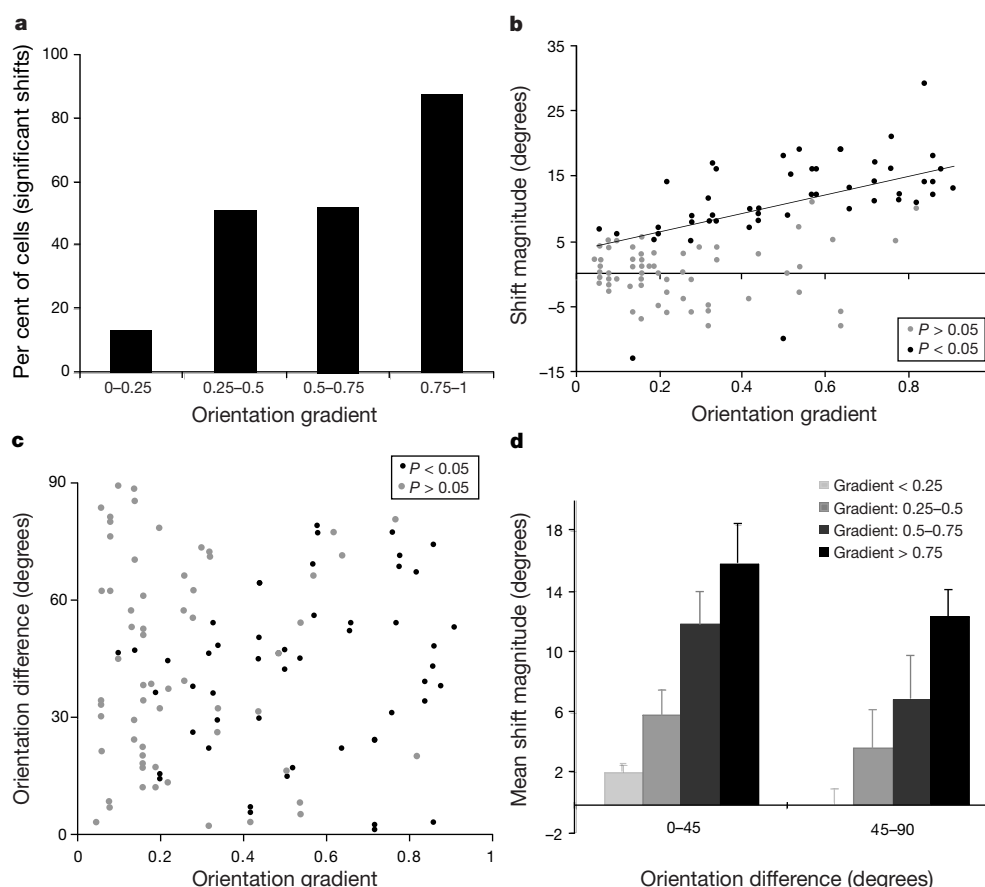


Figure 2 Population analysis of adaptation-induced orientation plasticity. **a**, Fraction of cells in discrete orientation gradient ranges showing significant shifts in orientation preference after adaptation. **b**, Scatter plot ($n = 118$ cells) showing magnitude of the post-adaptation shift in preferred orientation (positive and negative numbers indicate repulsive and attractive shifts, respectively) as a function of the orientation gradient at the recording site. Cells that show significant shifts in preferred orientation on the basis of a trial-by-trial comparison during adaptation and control conditions are indicated in black ($P < 0.05$); those that do not show significant shifts are indicated in grey ($P > 0.05$). The

line represents the linear regression fit for the black data points. Of these cells, 86% (43 of 50 cells) recovered their control orientation preference ($P > 0.05$, Student's t -test; comparing preferred orientations during control and after recovery). **c**, Scatter plot of the absolute difference ($\Delta\theta$) between the adapting orientation and the control preferred orientation as a function of orientation gradient at the recording site. **d**, Mean shift magnitude in orientation gradient bins for low and high $\Delta\theta$ ($\Delta\theta < 45^\circ$ and $\Delta\theta > 45^\circ$). Error bars indicate s.e.m.

gradient regions (Fig. 2d).

It is possible that the larger changes in orientation tuning observed at the high-gradient recording sites are due to inhomogeneities in the response properties of neurons across the cortical surface. For instance, higher firing rates or larger tuning bandwidths of neurons in regions of high orientation gradient could possibly cause larger shifts in orientation tuning. We therefore examined the relationship between the orientation gradient at each recording site and both the pre-adaptation peak firing rate and the orientation tuning width. We found that the peak response at the optimal orientation and the orientation tuning width (estimated by calculating the orientation selectivity index (OSI); see Methods) of the neurons in our population are not correlated with the orientation

gradient (peak response correlation coefficient $r = -0.09$; OSI correlation coefficient $r = -0.05$; $P > 0.1$ for each comparison); thus these properties are homogeneously represented across the cortical surface (see also ref. 9).

To understand the differences in adaptation-induced orientation plasticity between pinwheel centres and iso-orientation domains, we examined quantitatively the relationship between the local orientation distribution at the recording site and the degree of plasticity. On the basis of anatomical and physiological data in V1 (refs 8, 19–21), we postulated that local excitatory and inhibitory inputs to cortical cells originate from within a radius of about 500 μm around the cell body, and that adaptation is the result of integration of local inputs from a pool of neurons within this

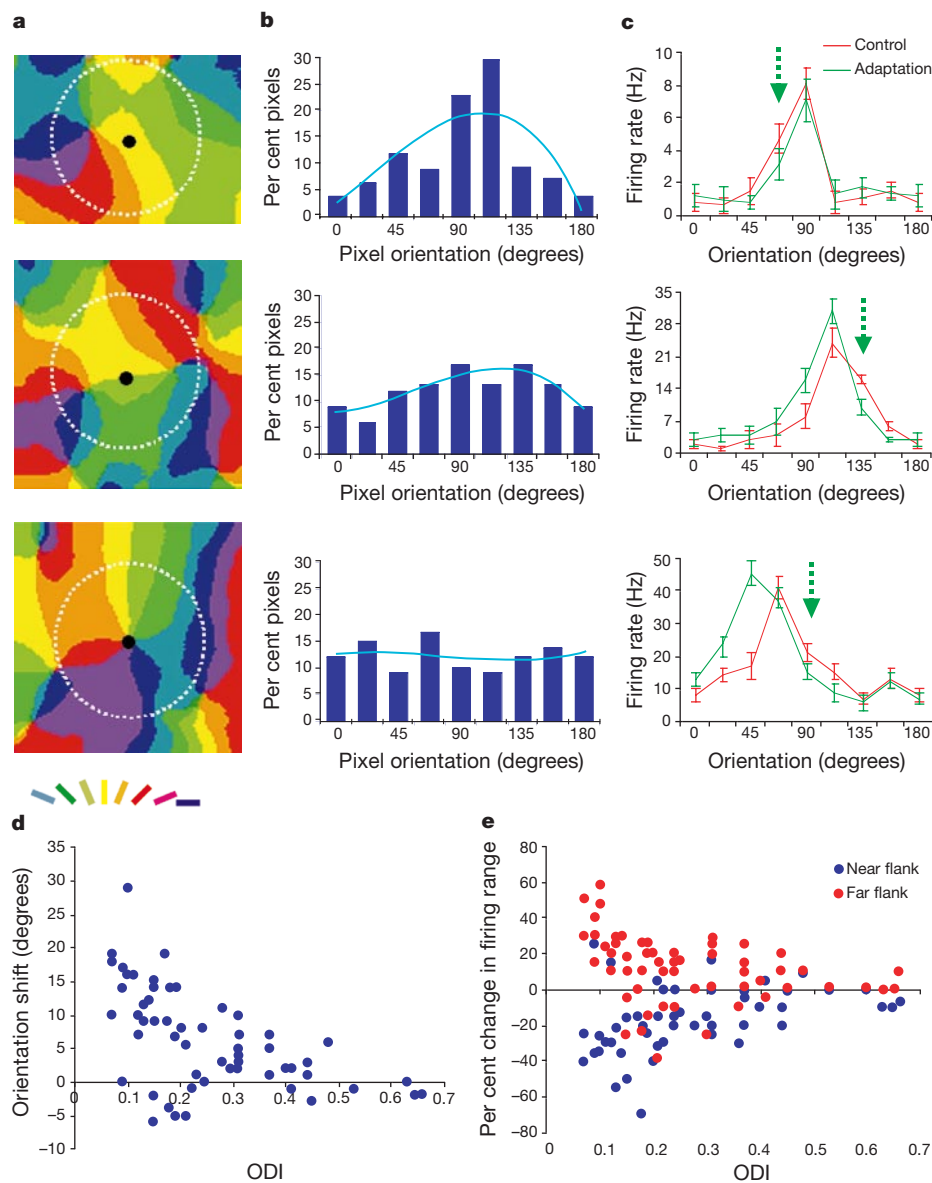


Figure 3 Relationship between local orientation distribution at the recording site and the degree of orientation plasticity. **a**, Three recording sites (filled circles) placed within an iso-orientation domain (top; gradient 0.06), between two pinwheel centres with input from neighbouring domains (middle; gradient 0.32), and in a pinwheel centre (bottom; gradient 0.64). The basin of local intracortical inputs is indicated by the white dotted circle (radius 500 μm). **b**, The percentage of pixels at each orientation within 500 μm of the recording site, after pooling the pixels into eight orientation bins between 0° and 180°. The light blue curve represents the best third-order polynomial fit to each histogram. **c**, Orientation

tuning curves during control (blue) and adaptation (green) for the cells recorded at the locations shown in **a**. The adapting orientation is marked by the green arrow. Orientation shifts for the three cells after adaptation are: top (6.8°), middle (11.4°) and bottom (19°). Each graph represents mean values $\pm$ s.e.m. **d**, Scatter plot ($n = 59$ cells) showing magnitude of the post-adaptation shift in preferred orientation as a function of the orientation distribution index (ODI) at the recording site. **e**, Scatter plot ($n = 59$ cells) showing the change in firing rate of responses on the near and far flanks of the orientation tuning curve as a function of the ODI at the recording site.

region. When the recording site is in the middle of an iso-orientation domain (Fig. 3a, top), neurons within a 500- μm radius have a preponderance of orientation preferences similar to the recorded neuron (Fig. 3b, top), whereas when the recording site is near a pinwheel centre (Fig. 3a, bottom), local inputs arise from domains of all orientations (Fig. 3b, bottom). Neurons at intermediate locations have intermediate orientation distribution profiles (Fig. 3a–b, middle). The orientation distribution of inputs to a neuron is a qualitative predictor of the degree of change in its orientation preference. When it has a peaked profile, adaptation induces only small changes in orientation selectivity (Fig. 3c, top), whereas the orientation tuning curve undergoes a large change when the orientation distribution is flat (Fig. 3c, bottom).

We quantified the orientation spread of inputs at different locations in the map by calculating an orientation distribution index (ODI), which is a measure of the orientation tuning strength of local inputs to the recorded cell (see Methods). We divided our population of neurons into different $\Delta\theta$ ranges, and show here the data from 59 neurons that were located at least 500 μm from the imaged area borders and were studied with small $\Delta\theta$ (between 0–45°, for which orientation shifts are maximal (Fig. 2d); results from neurons with large $\Delta\theta$ are similar). We calculated the shift in preferred orientation and the change in firing rate on both flanks of the tuning curve as a function of the ODI at the recording site. Figure 3d shows an inverse relationship between the orientation shift magnitude and the ODI ($r = -0.56$, $P < 0.001$). Furthermore, as the orientation shifts are characterized by suppression^{14,18} of responses on the flank of the tuning curve towards the adapting

stimulus (near flank) and by facilitation¹⁴ of responses on the flank of the tuning curve away from the adapting stimulus (far flank), we examined how the structure of the orientation tuning curve changes depending on cortical location. Figure 3e shows that the magnitude of the changes in response on both flanks of the tuning curve is related to the ODI (near flank: $r = 0.33$; far flank: $r = -0.34$; $P < 0.01$ for each comparison). Thus, broader orientation distributions of intracortical inputs to V1 cells result in larger changes in orientation tuning and firing rate. Because low ODIs characterize mainly the neurons in high-gradient regions (correlation between orientation gradient and ODI, $r = -0.5$, $P < 0.001$), the foci of adaptation-induced orientation plasticity are closely related to the functional map of orientation preference in V1.

As pinwheel centres are related to peaks of ocular dominance^{22,23}, our results may reflect a relationship between adaptation effects and monocular regions. We therefore recorded neurons of different eye preference located in targeted regions of the orientation map (Fig. 4a). A further 38 cells were first stimulated monocularly to determine their ocular dominance, followed by a full adaptation protocol using binocular stimulation. For each cell, we calculated an eye preference index as the absolute difference between responses to the preferred orientation measured independently for each eye, divided by the sum of optimal responses to each eye. Figure 4b shows that there is no significant correlation between the adaptation-induced change in preferred orientation and the eye preference index ($r = 0.18$, $P > 0.1$). Although the shift in orientation preference is correlated with the orientation gradient at the recording site ($r = 0.41$, $P < 0.01$), dividing the population of cells into two

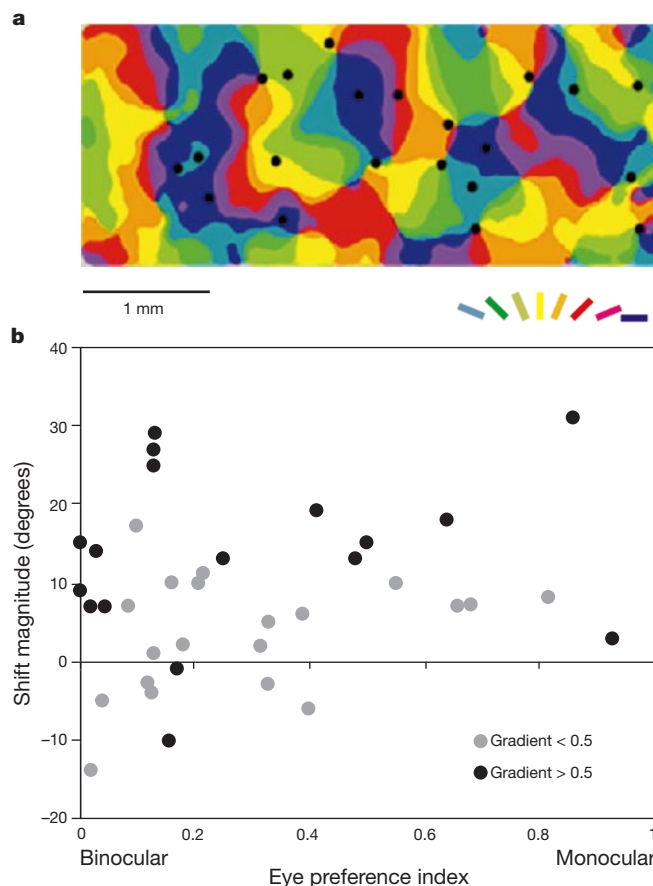


Figure 4 Relationship between eye preference and the degree of orientation plasticity. **a**, Orientation preference map; filled circles indicate the locations of representative targeted recording sites in one animal. **b**, Scatter plot ($n = 38$ cells) showing the magnitude of the post-adaptation shift in preferred orientation as a function of the eye

preference index and orientation gradient at the recording site. Grey, gradient < 0.5 ; black, gradient > 0.5 . Cells with an eye preference index of 0 are driven equally well by either eye, whereas those with an index of 1 are driven solely by one eye.

groups depending on the gradient magnitude also fails to yield a significant correlation with the eye preference index. However, as shown earlier (Fig. 2), the shift magnitude is larger for neurons located at higher gradient regions regardless of eye preference ($P < 0.002$, Student's t -test). Finally, of the population of 38 cells, 24 were studied with small $\Delta\theta$ ($< 45^\circ$). For these cells, we found a significant correlation between the shift magnitude and the orientation gradient ($r = 0.52$, $P < 0.01$), but no correlation between the shift magnitude and the eye preference index ($r = 0.11$, $P > 0.1$). Thus, the anisotropy in adaptation-induced plasticity of orientation tuning is unrelated to the ocular dominance of neurons.

The changes in orientation selectivity after adaptation reported here indicate a network mechanism that reorganizes responses across a broad range of orientations. At the level of single cells, hyperpolarization of neurons at or close to the adapting orientation (due to slow hyperpolarizing Ca^{2+} - and Na^+ -activated potassium channels²⁵ or to synaptic depression²⁶) can cause suppression of responses on the near flank of the tuning curve. Facilitation of responses on the far flank, however, requires disinhibition²⁷ and possibly amplification by means of local excitatory interactions²⁸ in the cortex. The strength of these effects, which determines the magnitude of the change in preferred orientation, depends on the location of neurons in the orientation map. Neurons in an iso-orientation domain would be only weakly activated by intracortical inputs with orientations that differ from the domain's preferred orientation, whereas neurons located at or near pinwheel centres would receive strong local inputs from neurons of all orientations. Therefore, altering the efficacy of these inputs through adaptation is likely to induce more profound changes in the orientation preference of neurons at or near pinwheel centres. This suggests that adaptation-induced orientation plasticity in V1 is an emergent property of a local cortical network overlaid on a non-uniform orientation map. These data indicate the existence of a map of orientation plasticity, closely related to the map of orientation preference, in which pinwheel centres constitute foci of plasticity and the orientation gradient is a measure of the degree of plasticity across V1. One question arising from this is whether there are similar plasticity maps related to other functional maps, such as those for eye preference, spatial frequency, or direction^{22–24}. There are several functional roles of a map of orientation plasticity. The map would allow dynamic short-term influences on neuronal responses at some locations while retaining invariant responses at other locations that would maintain a stable frame of reference. Specifically, preferential locations for adaptive changes may be a strategy that the visual cortex uses in the face of adaptation to the statistics of natural images²⁹ to prevent a global change in the orientation preference of visual cortical neurons. □

Methods

Animals

Adult cats were prepared for acute experiments according to protocols that were approved by MIT's Animal Care and Use Committee and conformed to NIH guidelines. Craniotomy followed by durotomy was performed to expose V1. Contact lenses were used to focus the eyes on a computer monitor.

Optical imaging

Techniques for intrinsic signal imaging were similar to those described^{6,14}. The orientation gradient map was obtained by applying a two-dimensional gradient operator to the orientation preference map. For each pixel, the spatial gradient was given by $\sqrt{dx^2 + dy^2}$, where $dx = |\theta_{x+1,y} - \theta_{x,y}|$ and $dy = |\theta_{x,y+1} - \theta_{x,y}|$ ($\theta_{x,y}$ is the preferred orientation of pixel (x,y) in the composite map). Values of dx and dy greater than 90° were subtracted from 180° , such that the maximum difference in preferred orientation was 90° . We related the gradient value at the recording site to changes in the orientation tuning of neurons by calculating the local orientation gradient as the mean of gradient values in a 3×3 pixel array centred at the recording location. The gradient values were normalized for analysis.

Electrophysiology

Stimuli consisted of 16 drifting square-wave gratings at high contrast, presented at orientations 22.5° apart at two directions of movement (opposite directions orthogonal to stimulus orientation). Typical stimulus parameters for V1 were: spatial frequency 0.5 cycle

per degree, temporal frequency 1 Hz. All stimuli were randomly interleaved. Stimuli were presented binocularly (except for determination of eye preference) and were shown to the animal on a 17-inch monitor positioned 30 cm in front of it. We recorded responses during three conditions: (1) before adaptation (control), when 16 drifting gratings were presented for 10 trials each for a total of 160 trials; (2) after 2 min of continuous adaptation to one grating of fixed orientation, when the 16 gratings presented for 7 trials each for a total of 112 trials were preceded by a 5-s 'topping-up' presentation of the adapting orientation to maintain the effects of previous adaptation; (3) after 10 min of recovery using a full-field uniform stimulus, when 16 gratings were presented in identical conditions as in the control condition. In all three conditions, the 16 gratings were presented for 2.5 s each. The full protocol, including control, adaptation and recovery periods, lasted about 2 h. To ensure stable recordings, individual cells were isolated using a spike-sort module (DataWave Technologies, v 5.0) that allowed the identification and discrimination of waveforms on the basis of their individual characteristics. We recorded at cortical depths between 300 and 800 μm from cells with initial orientation preferences covering the entire orientation range (between 0 and 180°).

The preferred orientation of neurons was calculated as described³⁰. The Fourier components were extracted for the orientation tuning curve and then normalized by dividing by the mean firing rate of the cell during stimulus presentation: $a = \sum_{i=0}^{n-1} R(\theta_i) \cos(2\theta_i)$; $b = \sum_{i=0}^{n-1} R(\theta_i) \sin(2\theta_i)$, where responses, $R(\theta_i)$, are obtained for a set of n test orientations θ_i , $i = 0, 1, \dots, n-1$, which are uniformly distributed over 0 to 180° . Preferred orientation, θ , was calculated as $\theta = 0.5 \arctan(b/a)$ if $a > 0$, or $\theta = 90 + 0.5 \arctan(b/a)$ if $a < 0$. If $a > 0$ and $b < 0$, $\theta = 180 + 0.5 \arctan(b/a)$. The OSI, which measures the strength of orientation tuning, was given by $c(M^{-1})$, where $c = \sqrt{a^2 + b^2}$, and M is the mean response magnitude averaged over all orientations. The ODI used in Fig. 3d–e was identical to the OSI, except that the 'orientation responses' consisted of the percentage of pixels of orientation θ_i around a recording site, pooled into eight orientation bins uniformly distributed over 0 to 180° . The Pearson test was used for all correlation comparisons.

Received 6 November 2000; accepted 30 January 2001.

- Gilbert, C. D. & Wiesel, T. N. Receptive field dynamics in adult primary visual cortex. *Nature* **356**, 150–152 (1992).
- Wang, X., Merzenich, M. M., Sameshima, K. & Jenkins, W. M. Remodelling of hand representation in adult cortex determined by timing of tactile stimulation. *Nature* **378**, 71–75 (1995).
- Fregnac, Y. & Shulz, D. E. Activity-dependent regulation of receptive field properties of cat area 17 by supervised Hebbian learning. *J. Neurobiol.* **41**, 69–82 (1999).
- Eysel, U. T. & Schweigart, G. Increased receptive field size in the surround of chronic lesions in the adult cat visual cortex. *Cereb. Cortex* **9**, 101–109 (1999).
- Hubel, D. H. & Wiesel, T. N. Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J. Comp. Neurol.* **158**, 267–293 (1974).
- Bonhoeffer, T. & Grünwald, A. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature* **353**, 429–431 (1991).
- Blasdel, G. G. Differential imaging of ocular dominance and orientation selectivity in monkey striate cortex. *J. Neurosci.* **12**, 3115–3138 (1992).
- Das, A. & Gilbert, C. D. Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature* **399**, 655–661 (1999).
- Maldonado, P. E., Godecke, I., Gray, C. M. & Bonhoeffer, T. Orientation selectivity in pinwheel centers in cat striate cortex. *Science* **276**, 1551–1555 (1997).
- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154 (1962).
- Reid, R. C. & Alonso, J. M. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature* **378**, 281–284 (1995).
- Ferster, D., Chung, S. & Wheat, H. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* **21**, 249–252 (1996).
- Somers, D. C., Nelson, S. B. & Sur, M. An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.* **15**, 5448–5465 (1995).
- Dragoi, V., Sharma, J. & Sur, M. Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron* **28**, 287–298 (2000).
- Hammond, P., Pomfrett, C. J. & Ahmed, B. Neural motion after-effects in the cat's striate cortex: orientation selectivity. *Vision Res.* **29**, 1671–1683 (1989).
- Saul, A. B. & Cynader, M. S. Adaptation in single units in visual cortex: the tuning of aftereffects in the spatial domain. *Vis. Neurosci.* **2**, 593–607 (1989).
- Dragoi, V., Sharma, J., Miller, E. K. M. & Sur, M. Dynamics of orientation adaptation in awake monkey V1 revealed by reverse correlation. *Soc. Neurosci. Abstr.* **25**, 1548 (1999).
- Muller, J. R., Metha, A. B., Krauskopf, J. & Lennie, P. Rapid adaptation in visual cortex to the structure of images. *Science* **285**, 1405–1408 (1999).
- Hata, Y., Tsumoto, T., Sato, H. & Tamura, H. Horizontal interactions between visual cortical neurones studied by cross-correlation analysis in the cat. *J. Physiol.* **441**, 593–614 (1991).
- Weliky, M., Kandler, K., Fitzpatrick, D. & Katz, L. C. Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. *Neuron* **15**, 541–552 (1995).
- Kisvarday, Z. F., Toth, E., Rausch, M. & Eysel, U. T. Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. *Cereb. Cortex* **7**, 605–618 (1997).
- Crair, M. C., Ruthazer, E. S., Gillespie, D. C. & Stryker, M. P. Ocular dominance peaks at pinwheel center singularities of the orientation map in cat visual cortex. *J. Neurophysiol.* **77**, 3381–3385 (1997).
- Hubener, M., Shoham, D., Grünwald, A. & Bonhoeffer, T. Spatial relationships among three columnar systems in cat area 17. *J. Neurosci.* **17**, 9270–9284 (1997).
- Isa, N. P., Trepel, C. & Stryker, M. P. Spatial frequency maps in cat visual cortex. *J. Neurosci.* **20**, 8504–8514 (2000).
- Sanchez-Vives, M. V., Nowak, L. G. & McCormick, D. A. Membrane mechanisms underlying contrast adaptation in cat area 17 *in vivo*. *J. Neurosci.* **20**, 4267–4285 (2000).
- Abbott, L. F., Varela, J. A., Sen, K. & Nelson, S. B. Synaptic depression and cortical gain control. *Science*

275, 220–224 (1997).

27. Dragoi, V. & Sur, M. Dynamic properties of recurrent inhibition in primary visual cortex: contrast and orientation dependence of contextual effects. *J. Neurophysiol.* **83**, 1019–1030 (2000).
28. Somers, D. C., Dragoi, V. & Sur, M. in *Cerebral Cortex: The Cat Primary Visual Cortex*. (eds Peters, A. & Payne, B.) (Academic, San Diego, in the press).
29. Coppola, D. M., Purves, H. R., McCoy, A. N. & Purves, D. The distribution of oriented contours in the real world. *Proc. Natl Acad. Sci. USA* **95**, 4002–4006 (1998).
30. Wörgötter, F. & Eysel, U. T. Correlations between directional and orientational tuning of cells in cat striate cortex. *Exp. Brain Res.* **83**, 665–669 (1991).

Acknowledgements

We thank J. Schummers for discussions. This work was supported by a Merck Fellowship (V.D.) and by grants from the NIH (M.S.).

Correspondence and requests for materials should be addressed to V.D.
(e-mail: vdragoi@ai.mit.edu).

BDNF controls dopamine D₃ receptor expression and triggers behavioural sensitization

Olivier Guillain*, Jorge Diaz†, Patrick Carroll‡, Nathalie Griffon*, Jean-Charles Schwartz*† & Pierre Sokoloff*

* Unité de Neurobiologie et Pharmacologie Moléculaire, INSERM U 109, Centre Paul Broca, 2ter rue d'Alésia, 75014 Paris, France

† Laboratoire de Physiologie, Université René Descartes, Avenue de l'Observatoire, 75006 Paris, France

‡ Unité de Développement et Pathologie du Motoneurone Spinal, INSERM U 382, IBDM, Campus de Luminy, Case 907, 13288 Marseille CEDEX 09, France

Brain-derived neurotrophic factor (BDNF), like other neurotrophins, is a polypeptidic factor initially regarded to be responsible for neuron proliferation, differentiation and survival, through its uptake at nerve terminals and retrograde transport to the cell body¹. A more diverse role for BDNF has emerged progressively from observations showing that it is also transported anterogradely^{2,3}, is released on neuron depolarization⁴, and triggers rapid intracellular signals⁴ and action potentials in central neurons⁵. Here we report that BDNF elicits long-term neuronal adaptations by controlling the responsiveness of its target neurons to the important neurotransmitter, dopamine. Using lesions and gene-targeted mice lacking BDNF, we show that BDNF from dopamine neurons is responsible for inducing normal expression of the dopamine D₃ receptor in nucleus accumbens^{6–8} both during development and in adulthood. BDNF from corticostriatal neurons³ also induces behavioural sensitization, by triggering overexpression of the D₃ receptor in striatum of hemiparkinsonian rats⁹. Our results suggest that BDNF may be an important determinant of pathophysiological conditions such as drug addiction¹⁰, schizophrenia¹¹ or Parkinson's disease¹², in which D₃ receptor expression is abnormal.

The D₃ receptor is normally expressed in the nucleus accumbens⁶, particularly in its shell division (AccSh), where dopamine is released by neurons originating from the ventral tegmental area. Unilateral and extensive destruction of dopamine neurons by local infusion of 6-hydroxydopamine (6-OHDA) elicits downregulation of D₃ receptor expression in AccSh of the denervated side¹³ (Fig. 1a), by deprivation of a factor of these neurons that is transported anterogradely and released on dopamine neuron activity¹³. This factor is distinct both from dopamine, as dopamine depletion or receptor blockade does not affect D₃ receptor expression, and from its known co-transmitters neurotensin and cholecystokinin¹³. We determined whether this factor might be BDNF, as this neurotrophin is present

in dopamine neurons from the ventral tegmental area¹⁴.

BDNF immunoreactivity is prominent in AccSh of normal rats¹⁵, as were messenger RNAs for the D₃ receptor and TrkB, the receptor for BDNF (Fig. 1a, b). In AccSh, D₃ receptor and TrkB mRNAs largely colocalize (Fig. 1c); 73 ± 3% of neurons expressing D₃ receptor mRNA also expressed TrkB mRNA (mean ± s.e.m. of counts obtained from 163 neurons in 3 animals). In the 6-OHDA-lesioned side, D₃ receptor mRNA was downregulated in AccSh (–38 ± 5% compared with unlesioned side, *n* = 6; *P* < 0.01, Mann–Whitney *U*-test), whereas TrkB mRNA was upregulated in AccSh (+16 ± 2.5%; *P* < 0.05) and striatum (+36 ± 4%; *P* < 0.02), in agreement with previous observations^{13,16}. BDNF (20 ng) was locally infused above the nucleus accumbens of 6-OHDA-lesioned rats, at a time when degeneration of dopamine neurons was complete, as assessed 1 week earlier by the ability of the dopamine agonist apomorphine to produce contralateral rotations (data not shown). The local infusion of BDNF reversed the 6-OHDA-induced decrease in D₃ receptor gene expression (Fig. 1d), indicating that exogenous BDNF compensates for the loss of dopamine neurons.

We next examined the effect of a BDNF-null mutation¹⁷ on D₃ receptor expression in developing mice. In rats, D₃ receptor expression appears in AccSh during the first postnatal week co-incident with its innervation by dopamine neurons, indicating that the factor maintaining D₃ receptor expression in adulthood may also trigger this expression during development⁷. In wild-type *BDNF*^{+/+} mice, D₃ receptor binding and mRNA in AccSh increased sharply from postnatal (P) days P9–P14 to P17–P23, whereas in homozygous *BDNF*^{–/–} mice, D₃ receptor binding and mRNA were low at P9–P14, and did not increase at later stages (Fig. 2a–c). These results show that BDNF is required for the normal development of D₃ receptor expression in AccSh.

The BDNF gene mutation did not impair the early development of dopamine neurons¹⁸, nor their later development, as tyrosine hydroxylase, a marker of these neurons, was not significantly

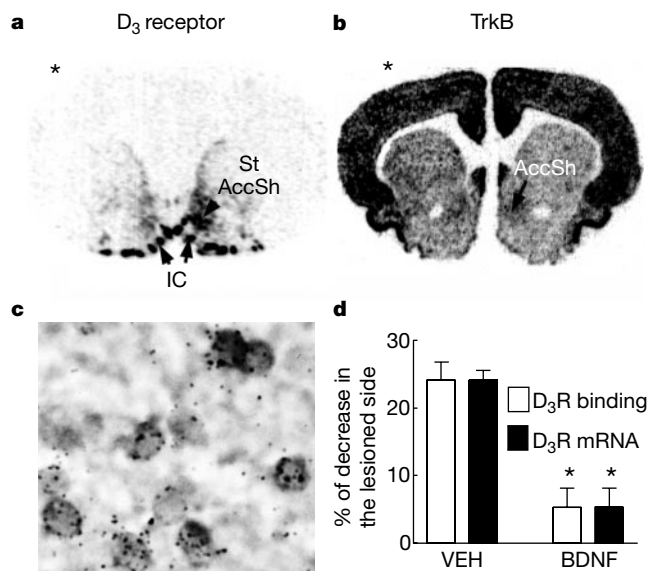


Figure 1 Dopamine D₃ receptor and TrkB expressions in 6-OHDA-lesioned rats.

a, b, Changes of *in situ* hybridization signals of D₃ receptor (D₃R; **a**) and TrkB (**b**) mRNAs on the lesioned side (asterisk). AccSh, shell of nucleus accumbens; IC, islands of Calleja; St, striatum. **c**, Colocalization of D₃ receptor (black dots) and TrkB (dark deposit) mRNAs in neurons of AccSh in control side. **d**, Effect of BDNF (20 ng per 4 μl), infused locally in both sides 21 d after the 6-OHDA lesion and 3 d before death, on the downregulation of D₃ receptor mRNA in AccSh. Results are expressed as per cent variation (mean ± s.e.m., *n* = 3–4) in the lesioned side compared with the control side. Asterisk, *P* < 0.05 versus saline-infused animals (VEH) by a Mann–Whitney *U*-test.

affected by the lack of BDNF in P9–P23 mice; tyrosine hydroxylase mRNA levels in the ventral tegmental area of *BDNF*^{−/−} mice were $89 \pm 8\%$ those of wild-type *BDNF*^{+/+} littermates ($P = 0.45$). This suggests that BDNF acts directly on neurons expressing D₃ receptor, rather than indirectly through an effect on the development of dopamine neurons. Moreover, BDNF deprivation selectively reduced the expression of D₃ receptor, and not that of dopamine D₁ and D₂ receptors, which are not, or only marginally down-regulated by 6-OHDA lesions⁹. Thus, D₁ and D₂ receptor mRNA levels in nucleus accumbens of *BDNF*^{−/−} mice were $108 \pm 6\%$ ($P = 0.62$) and $107 \pm 11\%$ ($P = 0.48$), respectively, of those of wild-type *BDNF*^{+/+} littermates.

In contrast, in the islands of Calleja TrkB is not expressed but the D₃ receptor is highly abundant (Fig. 1a, b); D₃ receptor expression occurs earlier during development⁷ and was not downregulated by the lesion¹³ (Fig. 1a), suggesting that it does not require BDNF. In accord, D₃ receptor binding in the islands of Calleja was not affected by the BDNF gene mutation ($F_{1,21} = 0.176$, $P = 0.68$, analysis of variance (ANOVA); see Fig. 2a), nor was D₃ receptor mRNA ($F_{1,21} = 0.102$, $P = 0.75$).

In unilaterally 6-OHDA-lesioned rats, repeated administration of levodopa, leading to extraneuronal dopamine formation, triggers D₃ receptor overexpression not only in AccSh but also in the denervated striatum—a brain structure in which D₃ receptor expression is hardly detectable⁹ (Figs 1 and 3). This D₃ receptor overexpression is responsible for the development of behavioural sensitization to levodopa—that is, a progressive enhancement of responsiveness, which appears as an increased number of levodopa-induced rotations (enhanced rotations are blocked by a preferential D₃ receptor antagonist⁹ and induced by a selective partial D₃

receptor agonist⁸).

During levodopa treatment of 6-OHDA-lesioned rats, we infused IgG–TrkB, a selective BDNF antagonist formed by fusion between the Fc-tail of human immunoglobulin-γ (IgG) and a part of TrkB¹⁹, into the denervated striatum by using osmotic minipumps; control animals received IgG. Diffusion of IgG or IgG–TrkB extended to the whole striatum, but not AccSh, as assessed using an anti-IgG antibody (data not shown). The treatment with IgG–TrkB impaired induction of both D₃ receptor expression (Fig. 3a, b) and behavioural sensitization (Fig. 3c). In addition, levodopa-induced induction of D₃ receptor protein was also impaired by treatment with IgG–TrkB; the ratios of D₃ receptor binding in the 6-OHDA-lesioned to control sides were 3.7 ± 0.7 , 1.5 ± 0.4 and 1.7 ± 0.4 in animals receiving IgG, TrkB–IgG ($0.05 \mu\text{g h}^{-1}$) or IgG–TrkB ($0.5 \mu\text{g h}^{-1}$), respectively ($n = 5-7$, $F_{2,14} = 5.62$, $P = 0.016$, ANOVA; $P < 0.05$ in animals treated with IgG–TrkB versus IgG).

Because striatal BDNF originates mainly from cortical neurons³, we examined the effects of a lesion of these neurons on levodopa-induced D₃ receptor expression and behavioural sensitization. In 6-OHDA-lesioned rats, we performed unilateral superficial cryo-lesions of the frontal cortex that were limited enough both to preserve striatal neurons, as assessed by a normal striatal substance P mRNA level in the damaged side (data not shown), and to leave initial levodopa-induced rotations unaffected (see the effect of the first levodopa administration in decorticated animals in Fig. 3e). Cortical ablation, however, partially impaired induction of D₃ receptor overexpression in striatum (ratios between D₃ mRNA levels in the 6-OHDA-lesioned and control sides were 3.7 ± 0.3 and 2.2 ± 0.6 in sham-operated and decorticated animals, respectively, $n = 3-5$, $P < 0.05$, see Fig. 3d) and behavioural sensitization (Fig. 3e), indicating that both processes require the participation of BDNF from corticostriatal neurons.

A single administration of levodopa progressively induced BDNF mRNA in the frontal, but not cingulate cortex in the 6-OHDA-lesioned side (Fig. 4a). BDNF mRNA was increased by levodopa mainly in cortical deeper layer V containing pyramidal cell bodies and layer VI (Fig. 4a); BDNF mRNA-positive cells in cortical layer V displayed the typical triangular shape of pyramidal cells (Fig. 4b). These cells project to various subcortical areas, notably including various striatal and accumbal areas²⁰, suggesting that levodopa may increase striatal BDNF release by cortico-striatal neurons³. Levodopa also induced BDNF mRNA in the control side, but this effect was much more limited in amplitude (Fig. 4c). This effect of levodopa on BDNF, like that on D₃ receptor overexpression and behavioural sensitization⁹, critically depends on the activation of a dopamine D₁ or D₅ receptor, as it was also produced by SKF 81297, a D₁/D₅ receptor agonist, but not by quinpirole, a D₂/D₃/D₄ receptor agonist (Fig. 4d). These results are consistent with the presence of D₁ receptors on cortical pyramidal cells²¹, and with the observation that stimulation of a D₁ or D₅ receptor under similar circumstances phosphorylates CRE-binding protein (ref. 22), a factor that activates BDNF gene transcription^{23,24}.

Because levodopa-induced D₃ receptor overexpression and behavioural sensitization require repeated treatments⁹, we next examined whether the changes in BDNF and its receptor expression persisted after repeated treatments with levodopa. Although repeated treatments induced tolerance to a subsequent levodopa administration, BDNF mRNA levels were still elevated ($+28 \pm 6\%$, $P < 0.05$ versus control side; data not shown) in the frontal cortex of the 6-OHDA-lesioned side of animals receiving 10 levodopa administrations. Moreover, TrkB overexpression in the 6-OHDA-lesioned side was accentuated after repeated administrations of levodopa (Fig. 5a), and mainly occurred in neurons expressing substance P mRNA (120/120 neurons expressing substance P mRNA also expressed TrkB mRNA; see Fig. 5b), in which D₃ receptor expression is induced by this treatment⁹. We conclude that induction of D₃ receptor expression in striatum is triggered by the combination of

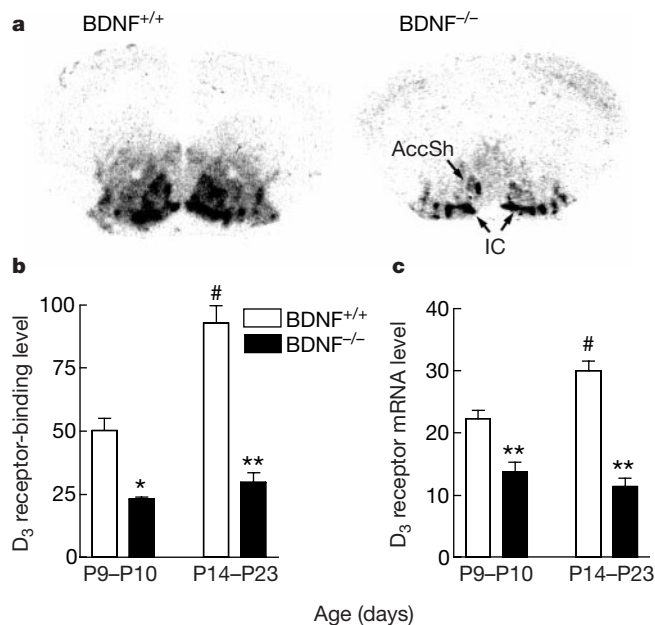


Figure 2 Impaired D₃ receptor expression in developing *BDNF*^{−/−} mice. **a**, Autoradiographic pictures of D₃ receptor binding, obtained with [¹²⁵I]7-OH-PIPAT, in wild-type *BDNF*^{+/+} or *BDNF*^{−/−} mice at postnatal day 23 (P23). **b**, Quantitative analysis of pictures in **a**, with animals at P9–P10 or P14–P23. The y axis units are in μCi per gram dry weight (also for **c**, and Figs 3b and 4d). Results are mean $\pm$ s.e.m. of values from 4–11 animals. There was a significant effect of genotype ($F_{1,21} = 40.52$, $P < 0.0001$), age ($F_{1,21} = 11.99$, $P = 0.0023$) and genotype $\times$ age interaction ($F_{1,21} = 6.41$, $P = 0.019$). **c**, Similar results were obtained on D₃ receptor mRNA levels, measured by quantitative *in situ* hybridization in the same animals, with a significant effect of genotype ($F_{1,21} = 50.96$, $P < 0.0001$), age ($F_{1,21} = 18.55$, $P = 0.0003$) and genotype $\times$ age interaction ($F_{1,21} = 19.08$, $P = 0.0003$). Asterisk, $P < 0.05$ and double asterisk, $P < 0.01$ versus *BDNF*^{+/+} littermates, hash, $P < 0.001$ versus wild-type P9–P10 mice.

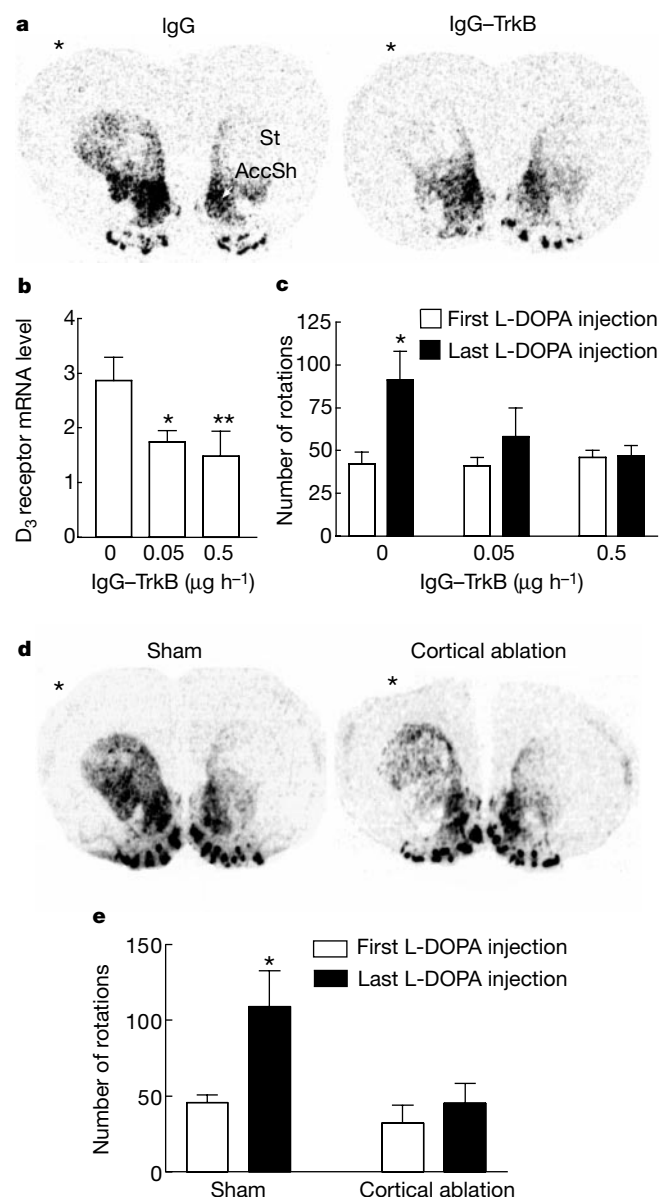


Figure 3 Levodopa-induced D₃ receptor expression and behavioural sensitization in 6-OHDA-lesioned rats require BDNF and participation of the frontal cortex. Asterisk denotes the lesioned side. **a**, *In situ* hybridization signals of D₃ receptor mRNA in animals receiving continuous infusions into the striatum of IgG (control animal, left) or IgG-TrkB (right) for 7 d and levodopa for 5 d. **b**, Quantitative analysis of pictures as in **a** showing a significant effect of treatment ($F_{2,14} = 7.42$, $P = 0.006$). Results are mean $\pm$ s.e.m. of values from 5–7 animals. Asterisk, $P < 0.05$ and double asterisk, $P < 0.01$ versus IgG-treated animals. **c**, Rotations were recorded after the first and last levodopa (L-DOPA) administrations in animals shown in **b**. There was a significant effect of the treatment ($F_{2,14} = 3.84$, $P = 0.046$). Asterisk, $P < 0.05$ versus first L-DOPA injection. **d**, *In situ* hybridization signals of D₃ receptor mRNA in animals without (left) or with (right) a partial cortical ablation, treated for 5 d with levodopa. **e**, Rotations were recorded in these animals after the first and last L-DOPA administrations. Results are mean $\pm$ s.e.m. of values from 3–5 animals. ($F_{3,11} = 7.31$, $P = 0.011$). Asterisk, $P < 0.05$ versus first L-DOPA administration in sham-operated animals.

a D₁/D₅ receptor stimulation-dependent elevation of BDNF in cortico-striatal neurons together with a denervation-dependent upregulation of striatal TrkB expression. Both processes are prominent in the 6-OHDA-lesioned side as compared with the control side, which accounts for the induction of D₃ receptor expression being restricted to the lesioned side. Moreover, BDNF-induced D₃

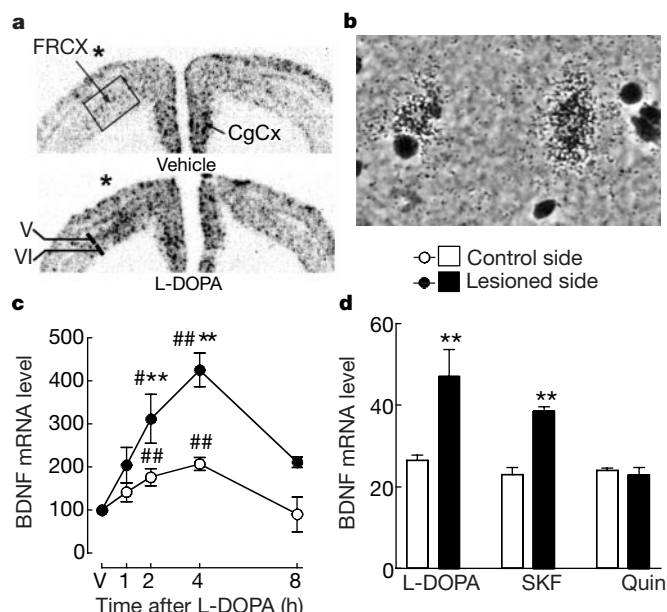


Figure 4 Levodopa upregulates BDNF expression in frontal cortex of 6-OHDA-lesioned rats. **a**, Increase in BDNF mRNA in the lesioned-side (asterisk) of frontal cortex (FRCX; boxed area 4 h after levodopa (L-DOPA)). CgCx, cingulate cortex. Roman numerals indicate cortical layers. **b**, Two BDNF mRNA-positive cells of deep layer 5 of FRCX of lesioned side of levodopa-treated animals. **c**, Time course of levodopa effects on BDNF mRNA levels in FRCX. Results are mean $\pm$ s.e.m. of values from 3–4 animals, expressed as a percentage of vehicle (V)-treated animals. There was a significant effect of treatment ($F_{1,24} = 34.89$, $P < 0.001$), time ($F_{4,24} = 21.68$, $P < 0.001$) and treatment $\times$ time interaction ($F_{4,24} = 5.37$, $P = 0.003$). Asterisk, $P < 0.05$ and double asterisk, $P < 0.01$ versus control, hash, $P < 0.05$ and double hash, $P < 0.01$ versus vehicle-treated animals. **d**, Compared effects of a single administration of levodopa, SKF 81297 (SKF) or quinpirole (Quin), on BDNF mRNA levels in FRCX. ($F_{2,14} = 10.89$, $P = 0.014$, double asterisk, $P < 0.01$ versus control side).

receptor expression causes a more pronounced disequilibrium in responsiveness to dopamine between the two sides, which leads to enhanced rotational behaviour.

Our observations, together with those suggesting the involvement of BDNF in long-term potentiation¹⁷, strengthen the view that BDNF elicits long-lasting changes in cerebral neurotransmission. This process, through which BDNF influences dopamine responsiveness, might be an important determinant in the aetiopathology and/or treatment of several conditions implicating dopamine. Thus, a progressive treatment-dependent elevation of corticostriatal BDNF in Parkinson's disease, like in 6-OHDA-lesioned rats, a reliable animal model, may account for the enhanced response to levodopa underlying its initial beneficial effect and leading to the abnormal movements often present in patients on long-term treatment. The same process might be critical in drug addiction. AccSh also receives innervations from areas highly enriched in BDNF¹⁵, such as hippocampus, lesions of which also downregulate D₃ receptor expression²⁵ presumably by a mechanism similar to that described here, and basolateral amygdala, from which BDNF is also transported anterogradely³. These brain areas, and notably the amygdala²⁶, process conditioned aspects of the environment, such as contextual cues associated with drug taking, the reactivity to which is controlled by the D₃ receptor⁸. We suggest that, through the above pathways and mechanisms, progressive changes in BDNF expression occurring during repeated drug-taking might alter dopamine responsiveness and induce drug-conditioned responses, a key process in drug addiction²⁷. In agreement with this, intra-accumbal infusions of BDNF enhance the behavioural sensitization to cocaine and response to cocaine-related stimuli in rats²⁸, and D₃

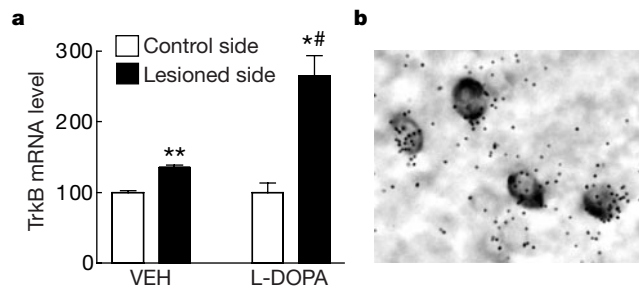


Figure 5 Levodopa treatments upregulate TrkB receptor expression in striatum of 6-OHDA-lesioned rats. **a**, TrkB mRNA in striatum of animals receiving repeated (twice a day for 5 d) administrations of vehicle (VEH) or levodopa (L-DOPA) expressed as percentage of control side in VEH-treated animals. Asterisk, $P < 0.05$ and double

asterisk, $P < 0.01$ versus control side; hash, $P < 0.05$ versus VEH-treated lesioned side. **b**, Colocalization of TrkB (black dots) and substance P (dark depot) mRNAs in striatal neurons after repeated administrations of levodopa.

receptor expression is elevated in accumbens of cocaine addicts¹⁰. More generally, through the feedback loops that AccSh forms with the prefrontal cortex, BDNF might affect the control of emotions and sensorimotor gating that is disturbed in some psychiatric disorders²⁹.

Methods

Animals, animal surgery and treatments

We generated mice carrying a deletion of the BDNF gene as described¹⁷. Wild-type and null-mutant mice were obtained by crossing heterozygotes, and all animals were genotyped by polymerase chain reaction. Wistar male rats (180–200 g; Iffa Credo) were anaesthetized with pentobarbital (50 mg per kg (body weight)) and infused with 6-OHDA (8 µg in 4 µl of 0.05% ascorbic acid in saline) in the medial forebrain bundle¹³. Three weeks later, the top half of the frontal cortex of some animals was exposed and cortical ablation was performed by applying liquid nitrogen for 1 min. Sham-operated animals had only anaesthesia and cortical exposure. BDNF (20 ng of human BDNF (Alamone Labs, Jerusalem, Israel) in 4 µl of 10 mM sodium phosphate buffer, pH 7.4) was bilaterally infused above the nucleus accumbens at coordinates: A, +1.7 mm; L, ±2.0 mm; H, –5.5 mm (Bregma). Human IgG or IgG–TrkB (Genentech) were infused through Alzet minipumps (model 2001) and cannulae, as described¹⁹, implanted in striatum at coordinates: A, +1.6 mm; L, ±2.5 mm; H, –5.2 mm (Bregma) 2 d before levodopa treatments. Single and repeated (twice daily) administrations of levodopa were performed with L-DOPA-methylester (50 mg per kg intraperitoneally (i.p.)), in combination with benserazide (12.5 mg per kg i.p.); some animals received quinpirole (1 mg per kg i.p.) or SKF 81297 (0.5 mg per kg i.p.) instead of levodopa. Levodopa was administered 3 weeks after the 6-OHDA lesion, and 4 d after sham-operation or cortical ablation, or 2 d after the start of IgG or IgG–TrkB treatment.

Receptor binding and *in situ* hybridization

Animals were killed and their brains frozen in isopentane and maintained at –30 °C. Slices (10 or 8 µm for mice) were either hybridized with ³³P-labelled antisense RNA (cRNA) probes or incubated with [¹²⁵I]R(+)-trans-7-hydroxy-2-[N-propyl-N-(3'-iodo-2'-propenyl)amino] tetralin ([¹²⁵I]trans-7-OH-PIPAT, 2,200 Ci mmol^{–1}; Amersham) as described¹³. cRNA probes for BDNF and TrkB corresponded to nucleotides 99–448 and 2,597–2,880, respectively; other probes have been described^{9,30}. In some experiments, a ³³P-labelled cRNA probe was combined with a digoxigenin-labelled cRNA probe, revealed by immunohistochemistry with the alkaline phosphatase reaction. Autoradiographic signals were quantified on 2–8 slices per animal using an image analyser (IMSTAR, Paris). Grey values were converted to µCi per gram dry weight using ¹⁴C or ³H standard stripes (Amersham).

Rotation behaviour

We measured rotations for 5 min, starting 35 min after administering levodopa following a 5-min period of habituation into vertical Plexiglas cylinders (30-cm diameter). Rotations were measured again after animals had received levodopa twice a day for 5 d.

Received 10 November 2000; accepted 25 January 2001.

1. Thoenen, H. Neurotrophins and neuronal plasticity. *Science* **270**, 593–598 (1995).
2. von Bartheld, C. S., Byers, M. R., Williams, R. & Bothwell, M. Anterograde transport of neurotrophins and axodendritic transfer in the developing visual system. *Nature* **379**, 830–833 (1996).
3. Altar, C. A. *et al.* Anterograde transport of brain-derived neurotrophic factor and its role in the brain. *Nature* **389**, 856–860 (1997).
4. Altar, C. A. & DiStefano, P. S. Neurotrophin trafficking by anterograde transport. *Trends Neurosci.* **21**, 433–437 (1998).
5. Kafitz, K. W., Rose, C. R., Thoenen, H. & Konnerth, A. Neurotrophin-evoked rapid excitation through TrkB receptors. *Nature* **401**, 918–921 (1999).
6. Sokoloff, P., Giros, B., Martres, M.-P., Bouthenet, M.-L. & Schwartz, J.-C. Molecular cloning and characterization of a novel dopamine receptor (D₃) as a target for neuroleptics. *Nature* **347**, 146–151 (1990).

7. Diaz, J. *et al.* Selective expression of dopamine D₃ receptor mRNA in proliferative zones during embryonic development of the rat brain. *J. Neurosci.* **17**, 4282–4292 (1997).
8. Pilla, M. *et al.* Selective inhibition of cocaine-seeking behaviour by a partial dopamine D₃ receptor agonist. *Nature* **400**, 371–375 (1999).
9. Bordet, R. *et al.* Induction of dopamine D₃ receptor as a mechanism of behavioral sensitization to levodopa. *Proc. Natl Acad. Sci. USA* **94**, 3363–3367 (1997).
10. Staley, J. K. & Mash, D. C. Adaptive increase in D₃ dopamine receptors in the brain reward circuits of human cocaine fatalities. *J. Neurosci.* **16**, 6100–6106 (1996).
11. Gurevich, E. V. *et al.* Mesolimbic dopamine D₃ receptors and use of antipsychotics in patients with schizophrenia. *Arch. Gen. Psychiatry* **54**, 225–232 (1997).
12. Ryoo, H. L., Pierrotti, D. & Joyce, J. N. Dopamine D₃ receptor is decreased and D₂ receptor is elevated in the striatum of Parkinson's disease. *Mov. Disord.* **13**, 788–797 (1998).
13. Lévesque, D. *et al.* A paradoxical regulation of the dopamine D₃ receptor expression suggests the involvement of an anterograde factor from dopamine neurons. *Proc. Natl Acad. Sci. USA* **92**, 1719–1723 (1995).
14. Seroogy, K. B. *et al.* Dopaminergic neurons in rat ventral midbrain express Brain-Derived Neurotrophic Factor and Neurotrophin-3 mRNAs. *J. Comp. Neurol.* **342**, 321–334 (1994).
15. Conner, J. & Lauter, J. C. Distribution of Brain-Derived Neurotrophic Factor (BDNF) protein and mRNA in the normal adult rat CNS: evidence for anterograde axonal transport. *J. Neurosci.* **17**, 2295–2313 (1997).
16. Numan, S. & Seroogy, K. B. Increased expression of TrkB mRNA in rat caudate-putamen following 6-OHDA lesions of the nigrostriatal pathway. *Eur. J. Neurosci.* **9**, 489–495 (1997).
17. Korte, M. *et al.* Hippocampal long-term potentiation is impaired in mice lacking brain-derived neurotrophic factor. *Proc. Natl Acad. Sci. USA* **92**, 8856–8860 (1995).
18. Ernfors, P., Lee, K.-F. & Jaenisch, R. Mice lacking brain-derived neurotrophic factor develop with sensory deficits. *Nature* **368**, 147–150 (1994).
19. Cabelli, R. J., Shelton, D. L., Segal, R. A. & Shatz, C. J. Blockade of endogenous ligands of TrkB inhibits formation of ocular dominance columns. *Neuron* **19**, 63–76 (1997).
20. Berendse, H. W., Galis-de-Graaf, Y. & Groenewegen, H. J. Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat. *J. Comp. Neurol.* **316**, 314–347 (1992).
21. Huang, Q. *et al.* Immunohistochemical localization of the D₁ dopamine receptor in rat brain reveals its axonal transport, pre- and post-synaptic localization, and prevalence in the basal ganglia, limbic system and thalamic reticular nucleus. *Proc. Natl Acad. Sci. USA* **89**, 11988–11992 (1992).
22. Cole, D. G., Kobierski, L. A., Konradi, C. & Hyman, S. E. 6-hydroxydopamine lesions of the rat substantia nigra up-regulate dopamine-induced phosphorylation of the cAMP-response element-binding protein in striatal neurons. *Proc. Natl Acad. Sci. USA* **91**, 9631–9635 (1994).
23. Tao, X., Finkbeiner, S., Arnold, D. B., Shaywitz, A. J. & Greenberg, M. E. Ca²⁺ influx regulates BDNF transcription by a CREB family transcription factor-dependent mechanism. *Neuron* **20**, 709–726 (1998).
24. Shieh, P. B., Hu, S. C., Bobb, K., Timmusk, T. & Ghosh, A. Identification of a signaling pathway involved in calcium regulation of BDNF expression. *Neuron* **20**, 727–740 (1998).
25. Flores, G., Barbeau, D., Quirion, R. & Srivastava, L. K. Decreased binding of dopamine D₃ receptors in limbic subregions after neonatal bilateral lesion of rat hippocampus. *J. Neurosci.* **16**, 2020–2026 (1996).
26. Whitelaw, R. B., Markou, A., Robbins, T. W. & Everitt, B. J. Excitotoxic lesions of the basolateral amygdala impair the acquisition of cocaine-seeking behaviour under a second-order schedule of reinforcement. *Psychopharmacology* **127**, 213–224 (1996).
27. O'Brien, C. P., Childress, A. R., McMellan, A. T. & Ehrman, R. A. A learning model of addiction. *Res. Publ. Assoc. Res. Nervous Mental Dis.* **70**, 157–177 (1992).
28. Horger, B. A. *et al.* Enhancement of locomotor activity and conditioned reward to cocaine by brain-derived neurotrophic factor. *J. Neurosci.* **19**, 4110–4122 (1999).
29. Carlsson, A. The current status of the dopamine hypothesis of schizophrenia. *Neuropsychopharmacology* **1**, 179–186 (1988).
30. Diaz, J. *et al.* Phenotypic characterization of neurons expressing the dopamine D₃ receptor. *Neuroscience* **65**, 731–745 (1995).

Acknowledgements

We thank V. Mignon for technical assistance and C. Degott for providing the anti-IgG antibody. O.G. received grants from Fondation pour la Recherche Médicale and Lundbeck Foundation. This work was supported by grants from the European Commission (Fifth Framework Programme) and the National Institute on Drug Abuse to P.S.

Correspondence and requests for reprints and materials should be addressed to P.S. (e-mail: sokol@broca.inserm.fr).

Membrane protein diffusion sets the speed of rod phototransduction

P. D. Calvert*, V. I. Govardovskii*, N. Krasnoperova†, R. E. Anderson‡, J. Lem† & C. L. Makino*

* Department of Ophthalmology, Harvard Medical School and the Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114, USA

† Departments of Ophthalmology and Genetics, Molecular Cardiology Research Institute, New England Medical Center and Tufts University School of Medicine, Boston, Massachusetts 02111, USA

‡ Departments of Ophthalmology and Cell Biology, Dean A McGee Eye Institute, University of Oklahoma Health Science Center, Oklahoma City, Oklahoma 73104, USA

Retinal rods signal the activation of a single receptor molecule by a photon¹. To ensure efficient photon capture, rods maintain about 10⁹ copies of rhodopsin densely packed into membranous disks². But a high packing density of rhodopsin may impede other steps in phototransduction that take place on the disk membrane³, by restricting the lateral movement of, and hence the rate of encounters between, the molecules involved^{4–6}. Although it has been suggested that lateral diffusion of proteins on the membrane sets the rate of onset of the photoresponse⁷, it was later argued that the subsequent processing of the complexes was the main determinant of this rate^{8,9}. The effects of protein density on response shut-off have not been reported. Here we show that a roughly 50% reduction in protein crowding achieved by the hemizygous knockout of rhodopsin in transgenic mice accelerates the rising phases and recoveries of flash responses by about 1.7-fold *in vivo*. Thus, in rods the rates of both response onset and recovery are set by the diffusional encounter frequency between proteins on the disk membrane.

Phototransduction begins when photoexcited rhodopsin (R*) and the G protein transducin (G) diffuse on the disk membrane, collide and then bind (reviewed in ref. 2). In this complex, R* catalyses the exchange of a GTP for a GDP on the Gα subunit and then the activated Gα (G*) is released. This process repeats many times for each R*, and constitutes the first amplifying step in the phototransduction cascade. G* diffuses along the membrane surface eventually activating a single cyclic GMP phosphodiesterase (PDE) subunit. Hydrolysis of cGMP by G*-activated PDE results in the closure of cationic channels in the plasma membrane and the hyperpolarizing photoresponse. Photoresponse recovery similarly requires the diffusional contact between proteins on the membrane (reviewed in ref. 10). R* shut-off begins with the phosphorylation of

rhodopsin by the membrane-associated rhodopsin kinase and is complete after the binding of a soluble arrestin molecule. The timely shut-off of G*-PDE depends on its interaction with a membrane-associated regulator of G-protein signalling, RGS9, which accelerates the hydrolysis of Gα-bound GTP. Finally, cGMP synthesized by guanylate cyclase reopens the cationic channels.

To identify which process—diffusional encounter between components or molecular processing—limits the speed of phototransduction, we accelerated the frequency of protein encounters in intact photoreceptors and looked for changes in photoresponse kinetics. The frequency of encounters between cascade components, ν_{enc} , is proportional to the sum of their diffusion coefficients¹¹. For a single R* and transducins²,

$$\nu_{\text{enc}} = 1.5(D_R + D_G)C_G \quad (1)$$

where D_R and D_G are the diffusion coefficients of R* and G, respectively and C_G is the packing density of transducin in the disk membrane. The diffusion coefficients are sensitive to molecular crowding, which reduces the mobilities of the proteins by decreasing their mean free paths of diffusion. This has been demonstrated by Monte Carlo simulations^{3,12} and by experimental measurements of the effects of protein concentrations on lateral diffusion^{4–6}.

For example, in the case of bacteriorhodopsin, an integral membrane protein whose cross-sectional area is similar to that of rhodopsin, the diffusion coefficient is inversely proportional to the protein packing density for molar phospholipid to protein ratios between 210 and 30 (ref. 4). In rod disk membranes, the phospholipid to rhodopsin ratio is about 65, and the diffusion coefficient for rhodopsin^{13,14} is consistent with the relationship described for bacteriorhodopsin. Thus, in the disk membrane, where diffusion of R* and transducins is restricted by an enormous number of quiescent rhodopsin molecules, ν_{enc} will increase with the removal of some of the rhodopsin molecules. This has been achieved by the deletion of one of the rhodopsin alleles in transgenic mice¹⁵.

Microspectrophotometric measurements showed that individual R^{+/-} rods contained about half of the normal amount of rhodopsin¹⁵. Other vital transduction components including transducin¹⁵, PDE¹⁵, RGS9, rhodopsin kinase, recoverin and arrestin (Fig. 1) were unchanged. The geometry of the outer segments and the spacing of the disks were also unchanged, suggesting that phospholipid molecules replaced the missing rhodopsins. We further support this conclusion with measurements of the phospholipid to rhodopsin ratio in purified disk membranes from R^{+/-} and wild-type rods (Table 1). Considering a density of 25,000–27,000 rhodopsins μm^{-2} in wild-type disk membranes² and a rhodopsin cross-sectional area of $\sim 10 \text{ nm}^2$ (ref. 16), rhodopsin normally occupies $\sim 27\%$ of the disk membrane surface area. The 2.7-fold increase in the phospholipid to rhodopsin ratio in R^{+/-} rods

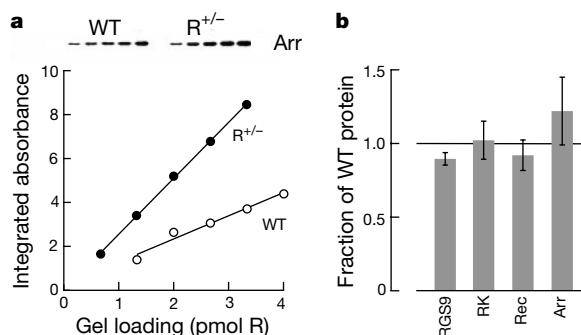


Figure 1 Normal levels of other proteins in R^{+/-} rod outer segments. **a**, Normal arrestin levels in R^{+/-} rods. Top, protein bands detected by ECL. Bottom, integrated optical density is proportional to the amount of rhodopsin loaded. Linear regression yielded slopes of 1.1 and 2.5 for wild type (WT; open circles) and R^{+/-} (filled circles), respectively, which are

consistent with a roughly 50% decrease in rhodopsin levels in R^{+/-} rods without a change in arrestin levels. **b**, Mean $\pm$ s.e.m. R^{+/-} protein levels relative to wild type, obtained from at least three determinations for each protein. Arr, arrestin; Rec, recoverin; RK, rhodopsin kinase.

indicated that rhodopsin density in the disks decreased about 2.2-fold. The addition of extra phospholipids to $R^{+/-}$ disks did not perturb the high unsaturated fatty acid content appreciably (Table 1), so the intrinsic membrane fluidity would not be expected to change. But the relief of membrane crowding would increase the diffusion coefficients of R^* , transducin and other membrane proteins roughly twofold^{3,4}. ν_{enc} and the rates of diffusion-limited reactions taking place on the membrane surface would increase proportionately.

For equivalent numbers of absorbed photons, flash responses in $R^{+/-}$ rods developed faster and peaked sooner than in wild-type rods, but their amplitudes were virtually equal (Fig. 2a, Table 1). We estimated the degree to which the onsets of responses to flashes suppressing less than 50% of the circulating current were acceler-

ated in $R^{+/-}$ rods by scaling the initial rising phases of wild-type responses to coincide with those of $R^{+/-}$ (see Fig. 2b). After averaging the scaling factors for the three dimmest flash strengths, the response onset was 1.67 ± 0.06 times faster in $R^{+/-}$ rods.

Bright flash responses were fitted with a quantitative model relating the slope of the rising phase to the degree of cascade amplification²:

$$\frac{r(t)}{r_{\max}} = 1 - \exp \left[-\frac{1}{2} \phi A (t - t_{\text{eff}})^2 \right] \quad (2)$$

where $r(t)$ is the response amplitude at time t after the flash, $r_{\max}$ is the maximal response, ϕ is the number of R^* , t_{eff} is an intrinsic delay in the response onset and A is the amplification constant. Equation (2) explicitly neglects inactivation processes that substantially

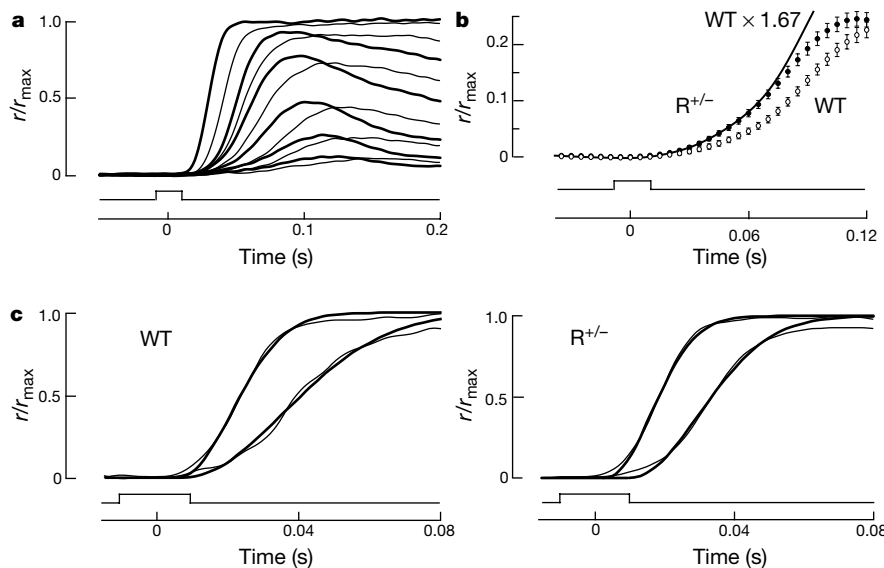


Figure 2 Accelerated flash response onset in $R^{+/-}$ rods. Lowest traces indicate the flash monitors. Numbers of R^* were based on photon-collecting areas of 0.47 and $0.23 \mu\text{m}^2$ for wild type (WT) and $R^{+/-}$, respectively. **a**, Flash responses from $R^{+/-}$ (thick traces) and WT (thin traces) rods. Each trace is the average of at least ten responses from six WT or nine $R^{+/-}$ rods except where indicated in parentheses. The 500-nm flash strengths (mean $\pm$ s.d.) were WT, 10.7 ± 1.5 , 22.1 ± 1.8 ($n = 5$), 40.4 ± 2.9 , 96.0 ± 6.9 , 172 ± 14 , $1,196 \pm 301$ ($n = 3$); $R^{+/-}$, 23.7 ± 2.8 , 44.5 ± 6.1 ($n = 7$), 101 ± 13 , 192 ± 26 , 404

± 17 ($n = 3$), $2,790 \pm 14$ photons μm^{-2} . **b**, Responses of WT (open circles) and $R^{+/-}$ (filled circles) rods in **a** to flashes resulting in $\sim 5.5 R^*$. Error bars indicated s.e.m. Solid line is WT response after scaling its amplitude by 1.67. **c**, Enhanced amplification constant in $R^{+/-}$ rods. Records were filtered at 80 Hz (see Methods). WT responses to 172 and 594 photons μm^{-2} (left) and $R^{+/-}$ responses to 400 and 1,400 photons μm^{-2} (right) were fitted with equation (2) (thick lines), yielding A values (in s^{-2}) of 15.2 and 16.3 (WT), and 23.4 and 22.4 ($R^{+/-}$).

Table 1 Flash response parameters and disk membrane composition

	Wild type	$R^{+/-}$	
Sensitivity, i_0 (photons μm^{-2}) [*]	60 ± 10 (14)	111 ± 10 (21)	$P < 0.001$
Dim flash response			
Time to peak (ms) [*]	140 ± 6 (10)	120 ± 2 (16)	$P < 0.01$
Integration time (ms) [*]	250 ± 18 (10)	180 ± 10 (16)	$P < 0.01$
Single-photon response amplitude (pA)	0.66 ± 0.08 (10)	0.64 ± 0.06 (10)	NS
Amplification constant, A (s^{-2})	13.6 ± 1.4	22.9 ± 1.5	$P < 0.01$
Saturating flash response			
τ_{c1} (ms)	200 ± 17 (10)	130 ± 7 (14)	$P < 0.001$
τ_{c2} (ms)	580 ± 92 (6)	350 ± 21 (11)	$P < 0.01$
$\text{Na}^+/\text{Ca}^{2+}, \text{K}^+$ exchanger			
τ_{ex} (ms)	105 ± 13 (9)	125 ± 22 (11)	NS
Magnitude (% of suppressible current)	7.7 ± 1.1 (9)	8.1 ± 1.0 (11)	NS
Disk membranes			
Phospholipid/rhodopsin (mol mol ⁻¹)	86 ± 12 (11)	236 ± 18 (14)	$P < 0.001$
Mean no. double bonds per fatty acid	2.7 ± 0.2 (11)	2.3 ± 0.1 (14)	NS
Saturated/unsaturated fatty acids (mol mol ⁻¹)	0.95 ± 0.12 (11)	1.07 ± 0.07 (14)	NS
Fatty acid length	19.4 ± 0.1 (11)	19.0 ± 0.1 (14)	NS

Values given as mean $\pm$ s.e.m. (number of determinations). i_0 is the half-saturating flash strength at 500 nm. Dim flash responses, whose amplitudes were less than a fifth of the maximum, had the same shape as the single-photon response and differed only in amplitude. Integration time was given by the integral of the response normalized by its amplitude. Amplitude of the single-photon response was found by dividing the response variance by the mean amplitude. A is the amplification constant from equation (2). The average A values were obtained from responses to flashes delivering 172, 594, 1,196 and 2,660 photons μm^{-2} (wild type) and 192, 404, 775, 1,400 photons μm^{-2} ($R^{+/-}$) as described in Fig. 2. τ_{c1} is the dominant time constant of the recovery for saturating flashes of $< 2,000$ photons μm^{-2} and τ_{c2} is that for saturating flashes of $> 2,000$ photons μm^{-2} . τ_{ex} is the time constant for $\text{Na}^+/\text{Ca}^{2+}, \text{K}^+$ exchange. NS, not significant.

^{*} Includes results from ref. 15.

influence the earliest detectable rise of responses to weak flashes but it is applicable to bright flash responses because their rising phases develop much more quickly than inactivation (Fig. 2c). In $R^{+/-}$ rods A was 1.7-fold higher than in wild-type rods (Table 1), in good agreement with the scaling factor found for dim flashes.

The amplification constant may be expressed in the following way²:

$$A = \nu_{RG} p_{GP} \beta n \quad (3)$$

where ν_{RG} is the rate of transducin activation by a single R^* , p_{GP} is the coupling efficiency between activated transducin and PDE, β is the hydrolytic rate constant of each PDE subunit and n is the Hill coefficient for activation of the cGMP-gated channel. As p_{GP} , β or n are not expected to have been altered by the deletion of a single rhodopsin allele, we attribute the entire 1.7-fold increase in A to a corresponding increase in ν_{RG} .

The rate of the multi-step activation of transducin may be expressed as:

$$\nu_{RG} = \frac{1}{\tau_{enc}/p_{RG} + \tau_{proc}} = \frac{1}{1/(\nu_{enc} p_{RG}) + \tau_{proc}} \quad (4)$$

Here $\tau_{enc} = 1/\nu_{enc}$ is the average time between R^* and G-GDP collisions, p_{RG} is the fraction of encounters that lead to formation of a complex, and the processing time τ_{proc} includes nucleotide exchange and dissociation of the complex. ν_{RG} must be lower than ν_{enc} as p_{RG} is not unity and processing time is finite. Estimates of ν_{enc} range from $7,000 \text{ s}^{-1}$ in amphibians to $17,000 \text{ s}^{-1}$ in mammals², whereas ν_{RG} is thought to be 1.4–50 times lower^{2,17}. The low values of ν_{RG} were attributed to a long τ_{proc} on the assumption that R^* -G coupling was highly efficient². However, in $R^{+/-}$ rods there is no reason to believe that the reduction of R packing density might shorten τ_{proc} of G activation. It is more likely that ν_{RG} increased because $1/(\nu_{enc} p_{RG})$ was slower than τ_{proc} . Then it is the frequency of productive encounters between R^* and G that sets the rate of photoresponse onset. Notably, the low rate of transducin activation reported¹⁷ infers that the fraction of effective collisions between R^* and G is 0.01–0.1, which is substantially lower than previously supposed.

Photoresponses also recovered more rapidly in $R^{+/-}$ rods. For all flash intensities studied, the substantially faster falling phase in the $R^{+/-}$ rods shortened the times to peak and abbreviated response

integration times (Fig. 3a, b; Table 1). Responses to bright flashes recovered in two phases. Only the faster phase was accelerated in $R^{+/-}$ rods. The slower phase, which probably arises from the failure of a small fraction of photoexcited rhodopsins to be phosphorylated by rhodopsin kinase¹⁸, was similar in $R^{+/-}$ and wild-type rods.

One possible explanation for accelerated response recovery in $R^{+/-}$ rods might be that the Ca^{2+} feedback systems operated more quickly. During the normal photoresponse the recovery is accelerated because of Ca^{2+} feedback¹⁰. Ca^{2+} enters the outer segment through the cGMP-gated channel and is extruded by a $\text{Na}^+/\text{Ca}^{2+}$, K^+ exchanger. Closure of the channels during the photoresponse blocks the entry of Ca^{2+} , but its efflux continues so intracellular Ca^{2+} declines. This decline is thought to relieve an inhibition of rhodopsin kinase by Ca^{2+} -recoverin, causing R^* to be shut-off faster, and to stimulate cGMP production through guanylate-cyclase-activating proteins.

Key factors that set the dynamic behaviour of Ca^{2+} feedback include the amplitude and rate of Ca^{2+} decline in response to light and the Ca^{2+} dependence of the feedback target¹⁹. Figure 4 shows that the amplitude and rate of light-stimulated Ca^{2+} decline were not different in wild-type and $R^{+/-}$ rods and that the Ca^{2+} dependence of cGMP synthesis, the most potent site of Ca^{2+} feedback in rods²⁰, was normal in $R^{+/-}$ rods (see also Table 1). Furthermore, the rhodopsin kinase and recoverin levels were normal (Fig. 1). Thus, Ca^{2+} feedback systems do not seem to be changed in $R^{+/-}$ rods.

Another explanation is that the cascade shuts off sooner owing to a reduction in protein crowding that accelerates protein–protein interactions. The shut-off of the phototransduction cascade has been studied using the Pepperberg method^{21,22}. Stimulation of rods with very bright flashes caused them to remain in saturation, with all cGMP-gated channels closed, for a period of time that depended on the flash strength. In some species the time in saturation was linearly related to the natural logarithm of flash intensity, where the slope gave the time constant of the slowest cascade shut-off step²¹. In mouse rods this relationship was not linear, indicating that several, progressively slower processes may limit the shut-off for flashes of increasing intensities (Fig. 3c). Nevertheless, the relationship between the time in saturation and the number of absorbed photons was less steep in the $R^{+/-}$ rods than in the wild-type rods. For comparison, the relations were approximated by linear regression of the data over two intensity domains. In each domain the slope for

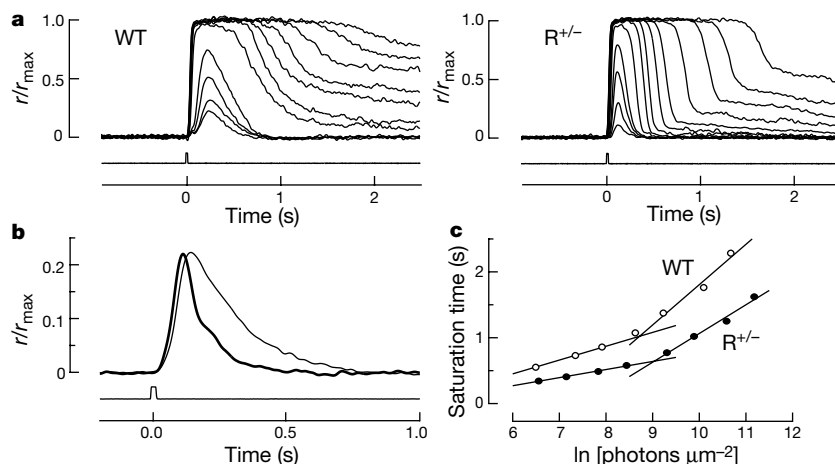


Figure 3 Accelerated flash response recovery in $R^{+/-}$ rods. **a**, Families of averaged flash responses from a wild-type (WT) and an $R^{+/-}$ rod. Flash strengths at 500 nm were (in photons μm^{-2}) WT, 6.4, 11.6, 23.2, 42.1, 659, 1,550, 2,810, 5,650, 10,200, 24,100 and 43,600; and $R^{+/-}$, 21, 38, 90, 164, 596, 1,400, 2,540, 5,100, 9,230, 21,800, 39,400, 79,200 and 143,000. The maximal response was 8.2 pA for WT and 11.3 pA for $R^{+/-}$. **b**, Averaged dim flash responses of five WT (thin trace) and seven $R^{+/-}$ (thick trace)

rods to $\sim 5.5 R^*$. **c**, Times that cells in **a** remained in saturation after bright flashes, measured from mid-flash until 20% recovery of the circulating current. The relationship for the $R^{+/-}$ rod was shifted left by 0.693 ln units to normalize for photon capture. Straight lines show linear regression analyses with slopes of 0.210 s for WT and 0.123 s for $R^{+/-}$ for flashes $< 2,000$ photons μm^{-2} , and 0.612 s for WT and 0.436 s for $R^{+/-}$ for flashes $> 2,000$ photons μm^{-2} .

$R^{+/-}$ rods was 1.5–1.7-fold less than that for the wild-type rods (Table 1). Thus, decreased membrane crowding also caused the dominant mechanism(s) controlling the cascade shut-off rate to proceed more rapidly.

On the one hand, the key rate-limiting recovery step may be rhodopsin quench, because in recoverin knockout mouse rods recovery was also accelerated²³. In these rods, R^* shut off sooner presumably because the higher concentration of Ca^{2+} -recoverin-free rhodopsin kinase increased the frequency of encounters between R^* and rhodopsin kinase. On the other hand, G^* -PDE shut-off has been implicated as the slowest recovery process in amphibian rods²². Other data²⁴ are consistent with the idea that formation of the complex between RGS9-G β 5 and G^* limits the rate of G^* shut-off. So, the time required for phosphate transfer to R^* by rhodopsin kinase or for RGS9-G β 5 accelerated G^* GTPase must be substantially shorter than the time between productive encounters of corresponding components. Increased rates of diffusion on the membrane will therefore cause faster quench of R^* and/or G^* .

In conclusion, lowering the rhodopsin concentration in rod photoreceptor membranes decreases photon capture, thereby

reducing threshold sensitivity. At the same time the photoresponse speeds up, improving temporal resolution of the visual stimulus. Conversely, raising the rhodopsin packing density would increase light absorption, but would slow and eventually stop phototransduction owing to excessive protein crowding. Optimal solution of the problem may depend on the specific behavioural needs of the photoreceptor's owner. Furthermore, as the response speed is set by the frequency of protein–protein encounters, the photoreceptor might alter its sensitivity and temporal resolution to match long-term changes in ambient lighting by modulating the levels of certain transduction proteins. It will be interesting to see whether similar trade-offs exist between speed and sensitivity in other membrane-bound signal transduction systems. $\square$

Methods

ROS preparation for biochemical analyses

Rod outer segments (ROSs) were collected from 8–10-week-old mice under infrared illumination²⁵. Mice were dark adapted for a minimum of 12 h, anaesthetized with CO_2 and killed by cervical dislocation. Retinas were isolated into Locke's solution containing 6.7% Optiprep (Fisher), vortexed and large pieces were allowed to sediment. The supernatant was loaded onto a step gradient consisting of 6.7, 8.3 and 20% Optiprep. After centrifugation for 15 min at 1,425 g ROSs were collected from the 8.3–20% gradient interface. Optiprep was removed by several washings with Locke's solution containing 0.1 mM phenylmethylsulphonyl fluoride, 2 mM EDTA and 15 μ g ml⁻¹ each of aprotinin, leupeptin and pepstatin. Rhodopsin content was determined spectrophotometrically.

Determination of relative ROS protein levels

ROS samples from $R^{+/+}$ and $R^{+/-}$ animals were dissolved in SDS sample buffer and specific protein levels were determined by western blot. Four to five samples containing increasing amounts of R from each mouse genotype were run in neighbouring lanes on a single 12.5% polyacrylamide gel, and transferred onto nitrocellulose membrane. We probed blots with primary antibodies as described¹⁵ along with an additional anti-RGS9 antibody²⁶. Horseradish peroxidase-conjugated secondary antibodies were used at a dilution of 1:10,000, visualized by enhanced chemiluminescence (ECL; Pierce), and proteins were quantified by scanning densitometry using an AlphaImager (Alpha Innotech, San Leandro, CA). Integrated optical densities of bands were plotted as a function of the amount of rhodopsin loaded. We used the linear region of this relationship to determine the amount of protein in $R^{+/-}$ ROSs relative to that in wild type.

Determination of guanylate cyclase activity

The assay of guanylate cyclase activity was carried out essentially as described¹⁹. The reaction was started by adding 5 μ l of pseudo-intracellular medium containing GTP to 10 μ l of ROS suspension. Final concentrations of reactants were (in mM): 0.02 ($R^{+/+}$) or 0.01 ($R^{+/-}$), 2 Mg^{2+} , 0.15 zaprinast, 3 [³²P]GTP- α , 30 cGMP, 0.06 ATP and the Ca^{2+} /BAPTA ratio required for the desired free Ca^{2+} . We quenched the reaction after 2 min by adding 60 μ l 50 mM EDTA and boiling for 1 min. Samples were centrifuged for 10 min on a Beckman Microfuge E. Supernatant (5 μ l) was loaded onto PEI-cellulose thin layer chromatography plates (EM Sciences, Gibbstown, NJ). Nucleotide separation was carried out with 0.2 mM LiCl. cGMP was visualized with ultraviolet light, cut from the plate, eluted with 2.0 M LiCl and counted in 10 ml of ScintiSafe cocktail.

Photoreceptor lipid content

Lipid analysis was carried out on photoreceptor disks prepared from ROSs purified as stated above. The disk preparation was a modification of the Ficoll floatation method developed for bovine disks²⁷. We determined phospholipid content and fatty acid composition as described²⁸.

Single-cell suction electrode recording

Photoresponses were recorded as described²⁹. We carried out all procedures under infrared illumination. Dark-adapted mice were anaesthetized with CO_2 and killed by cervical dislocation. Retinas were isolated into Leibovitz's L-15 medium (Gibco). Residual vitreous humor was removed and the tissue stored on ice. Small pieces of tissue were chopped finely, treated with DNase (type IV; Sigma), transferred to an experimental chamber and perfused with an enriched Locke's solution saturated with 95% O_2 /5% CO_2 . Locke's solution contained (in mM): 144 Na^+ , 3.6 K^+ , 1.2 Ca^{2+} , 2.4 Mg^{2+} , 123.3 Cl^- , 10 HEPES, 20 HCO_3^- , 0.02 EDTA, 10 glucose, 0.5 glutamate, 3 succinate, BME vitamins, MEM amino acids; pH 7.4. A rod outer segment was pulled into a glass suction electrode and the membrane currents monitored with a current-to-voltage recorder (Axopatch 200A, Axon Instruments, CA) at 36–38 °C. The solution inside the electrode did not contain vitamins or amino acids, and HCO_3^- was replaced with Cl^- . We stimulated rods with 21-ms flashes from a xenon arc source passing through a six-cavity interference filter that transmitted 500-nm light with a half-maximal transmission bandwidth of 10 nm (Omega Optical, Brattleboro, VT). Records were low-pass filtered at 30 Hz (–3 dB, 8-pole Bessel filter) and digitized at 400 Hz. The delay introduced by low-pass filtering was not removed from

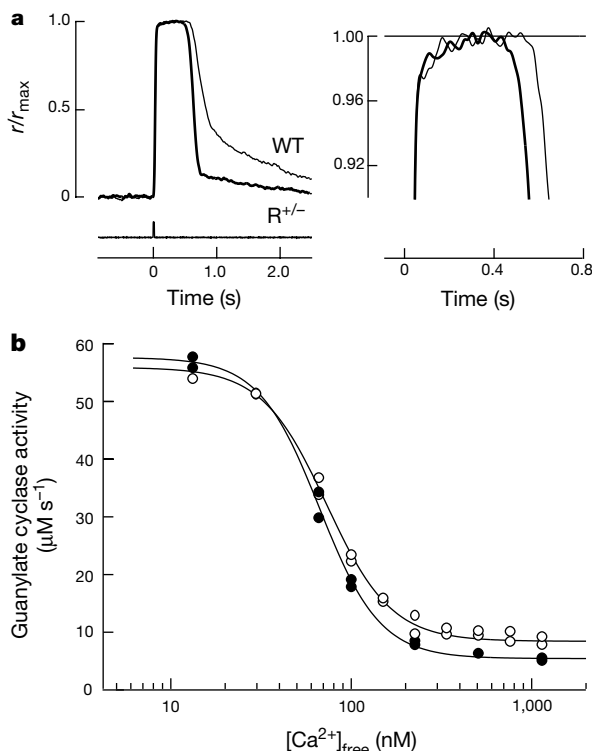


Figure 4 Ca^{2+} feedback was unperturbed in $R^{+/-}$ rods. **a**, Ca^{2+} dynamics in the rod outer segment. Left, saturating flashes prevented the entry of Ca^{2+} by closing all cGMP-gated channels in the plasma membrane. Right, the residual current at the peak of the response was due to removal of intracellular Ca^{2+} by the electrogenic $Na^+/Ca^{2+}, K^+$ exchanger³⁰ (expanded scale). The similar amplitude and kinetics between wild-type (WT) (thin traces) and $R^{+/-}$ rods (thick traces) indicates that the intracellular Ca^{2+} content in darkness and the rate of light-stimulated Ca^{2+} removal were unchanged in $R^{+/-}$ rods. **b**, Normal Ca^{2+} dependence of cGMP synthesis in $R^{+/-}$ rods. Rate of cGMP synthesis is expressed as the change in concentration in the outer segment cytoplasmic volume over time (μ M s⁻¹). Continuous lines show fits with the Hill equation: $\alpha(Ca) = \alpha_{max} - \Delta\alpha (Ca^n)/(Ca^n + k^n)$, where α is the rate of cGMP synthesis, α_{max} is the maximal rate, $\Delta\alpha$ is the difference between the highest and lowest rates, Ca is $[Ca^{2+}]_{free}$, k is a constant, n is the Hill coefficient. WT (open circles): $\alpha_{max} = 55.9 \mu$ M s⁻¹, $\Delta\alpha = 47.5 \mu$ M s⁻¹, $k = 73.0$ nM, $n = 2.4$. $R^{+/-}$ (filled circles): $\alpha_{max} = 57.6 \mu$ M s⁻¹, $\Delta\alpha = 52.2 \mu$ M s⁻¹, $k = 66.6$ nM, $n = 2.6$.

responses unless otherwise noted. Filtering at 30 Hz influenced the determination of the cell amplification constant from responses to bright flashes. Therefore, the filtering was digitally removed by deconvolution (Numerical Recipes) and then the records were low-pass filtered at 80 Hz.

Received 8 November 2000; accepted 31 January 2001.

1. Baylor, D. A., Lamb, T. D. & Yau, K.-W. Responses of retinal rods to single photons. *J. Physiol. (Lond.)* **288**, 613–634 (1979).
2. Pugh, E. N. Jr & Lamb, T. D. Amplification and kinetics of the activation steps in phototransduction. *Biochim. Biophys. Acta* **1141**, 111–149 (1993).
3. Saxton, M. J. & Owlicki, J. C. Concentration effects on reactions in membranes: rhodopsin and transducin. *Biochim. Biophys. Acta* **979**, 27–34 (1989).
4. Peters, R. & Cherry, R. J. Lateral and rotational diffusion of bacteriorhodopsin in lipid bilayers: experimental test of the Saffman–Delbruck equations. *Proc. Natl Acad. Sci. USA* **79**, 4317–4321 (1982).
5. Tank, D. W., Wu, E. S., Meers, P. R. & Webb, W. W. Lateral diffusion of gramicidin C in phospholipid multibilayers. Effects of cholesterol and high gramicidin concentration. *Biophys. J.* **40**, 129–135 (1982).
6. Tilton, R. D., Gast, A. P. & Robertson, C. R. Surface diffusion of interacting proteins. Effect of concentration on the lateral mobility of adsorbed bovine serum albumin. *Biophys. J.* **63**, 1321–1326 (1990).
7. Kahlert, M. & Hofmann, K. P. Reaction rate and collisional efficiency of the rhodopsin-transducin system in intact retinal rods. *Biophys. J.* **59**, 375–386 (1991).
8. Bruckert, F., Chabre, M. & Vuong, T. M. Kinetic analysis of the activation of transducin by photoexcited rhodopsin. Influence of the lateral diffusion of transducin and competition of guanosine diphosphate and guanosine triphosphate for the nucleotide site. *Biophys. J.* **63**, 616–629 (1992).
9. Heck, M. & Hofmann, K. P. G-protein-effector coupling: a real-time light-scattering assay for transducin–phosphodiesterase interaction. *Biochemistry* **32**, 8220–8227 (1993).
10. Pugh, E. N. Jr, Nikonov, S. & Lamb, T. D. Molecular mechanisms of vertebrate photoreceptor light adaptation. *Curr. Opin. Neurobiol.* **9**, 410–418 (1999).
11. Naqvi, K. R. Diffusion-controlled reactions in two-dimensional fluids: discussion of measurements of lateral diffusion of lipids in biological membranes. *Chem. Phys. Lett.* **28**, 280–284 (1974).
12. Pink, D. A. Protein lateral movement in lipid bilayers. Simulation studies of its dependence upon protein concentration. *Biochim. Biophys. Acta* **818**, 200–204 (1985).
13. Liebman, P. A. & Entine, G. Lateral diffusion of visual pigment in photoreceptor disk membranes. *Science* **185**, 457–459 (1974).
14. Poo, M. & Cone, R. A. Lateral diffusion of rhodopsin in the photoreceptor membrane. *Nature* **247**, 438–441 (1974).
15. Lem, J. *et al.* Morphological, physiological, and biochemical changes in rhodopsin knockout mice. *Proc. Natl Acad. Sci. USA* **96**, 736–741 (1999).
16. Palczewski, K. *et al.* Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739–745 (2000).
17. Leskov, I. B. *et al.* The gain of rod phototransduction: reconciliation of biochemical and electrophysiological measurements. *Neuron* **27**, 525–537 (2000).
18. Chen, J., Makino, C. L., Peachey, N. S., Baylor, D. A. & Simon, M. I. Mechanisms of rhodopsin inactivation *in vivo* as revealed by a COOH-terminal truncation mutant. *Science* **267**, 374–377 (1995).
19. Calvert, P. D., Ho, T. W., LeFebvre, Y. M. & Arshavsky, V. Y. Onset of feedback reactions underlying vertebrate rod photoreceptor light adaptation. *J. Gen. Physiol.* **111**, 39–51 (1998).
20. Koutalos, Y. & Yau, K.-W. Regulation of sensitivity in vertebrate rod photoreceptors by calcium. *Trends Neurosci.* **19**, 73–81 (1996).
21. Pepperberg, D. R. *et al.* Light-dependent delay in the falling phase of the retinal rod photoresponse. *Vis. Neurosci.* **8**, 9–18 (1992).
22. Nikonov, S., Engheta, H. & Pugh, E. N. Jr Kinetics of recovery of the dark-adapted salamander rod photoresponse. *J. Gen. Physiol.* **111**, 7–37 (1998).
23. Dodd, R. L., Makino, C. L., Chen, J., Simon, M. I. & Baylor, D. A. Visual transduction in transgenic mouse rods lacking recoverin. *Invest. Ophthalmol. Vis. Sci.* **36**, S641 (1995).
24. Skiba, N. P., Hopp, J. A. & Arshavsky, V. Y. The effector enzyme regulates the duration of G protein signaling in vertebrate photoreceptors by increasing the affinity between transducin and RGS protein. *J. Biol. Chem.* **275**, 32716–32720 (2000).
25. Tsang, S. H. *et al.* Role for the target enzyme in deactivation of photoreceptor G protein *in vivo*. *Science* **282**, 117–121 (1998).
26. Makino, E. R., Handy, J. W., Li, T. & Arshavsky, V. Y. The GTPase activating factor for transducin in rod photoreceptors is the complex between RGS9 and type 5 G protein β subunit. *Proc. Natl Acad. Sci. USA* **96**, 1947–1952 (1999).
27. Smith, H. G. Jr, Stubbs, G. W. & Litman, B. J. The isolation and purification of osmotically intact discs from retinal rod outer segments. *Exp. Eye Res.* **20**, 211–217 (1975).
28. Yuan, C., Chen, H., Anderson, R. E., Kuwata, O. & Ebrey, T. G. The unique lipid composition of gecko (Gekko Gekko) photoreceptor outer segment membranes. *Comp. Biochem. Physiol. B* **120**, 785–789 (1998).
29. Sung, C.-H., Makino, C., Baylor, D. & Nathans, J. A rhodopsin gene mutation responsible for autosomal dominant retinitis pigmentosa results in a protein that is defective in localization to the photoreceptor outer segment. *J. Neurosci.* **14**, 5818–5833 (1994).
30. Yau, K.-W. & Nakatani, K. Electrogenic Na–Ca exchange in retinal rod outer segment. *Nature* **311**, 661–663 (1984).

Acknowledgements

We thank P. Farabella, M. McClellan and M. Maude for technical assistance; R. Lefkowitz, A. Dizhoor, W. Smith, R. Lee, M. Simon, E. Makino and V. Arshavsky for antibodies; and E. Pugh Jr, D. Baylor and V. Arshavsky for comments on the manuscript. This work was supported by the E. Mathilda Ziegler Foundation (C.L.M.), Milton Fund (C.L.M.), the NIH (P.D.C., J.L., R.E.A.), Lion's of Massachusetts (C.L.M., J.L.), Fight for Sight (J.L.), Research to Prevent Blindness (C.L.M., J.L., R.E.A.), Foundation Fighting Blindness (J.L., R.E.A.), and the US Civilian Research and Development Foundation (V.I.G.).

Correspondence should be addressed to P.D.C. (e-mail: pdcalvert@meei.harvard.edu).

Translational repression determines a neuronal potential in *Drosophila* asymmetric cell division

Masataka Okabe^{*†}, Takao Imai^{†‡§}, Mitsuhiro Kurusu^{||§}, Yasushi Hiromi^{*} & Hideyuki Okano^{‡#}

^{*} Division of Developmental Genetics, National Institute of Genetics, and Department of Genetics, Graduate University for Advanced Studies, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

[‡] Division of Neuroanatomy (D12), Department of Neuroscience, Biomedical Research Center, Osaka University Graduate School of Medicine, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan, and Department of Physiology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku, Tokyo 160-8582, Japan

[§] Laboratory of Neuroscience, Division of Biophysical Engineering, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan

^{||} Department of Molecular Neurobiology, Institute of Basic Medical Sciences, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaragi 305-8572, Japan

[#] Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Corporation (JST), 2-6-15 Shibakoen, Minato-ku, Tokyo 105-0011, Japan

[†] These authors contributed equally to this work

Asymmetric cell division is a fundamental strategy for generating cellular diversity during animal development¹. Daughter cells manifest asymmetry in their differential gene expression. Transcriptional regulation of this process has been the focus of many studies, whereas cell-type-specific 'translational' regulation has been considered to have a more minor role. During sensory organ development in *Drosophila*, Notch signalling directs the asymmetry between neuronal and non-neuronal lineages², and a zinc-finger transcriptional repressor Tramtrack69 (TTK69) acts downstream of Notch as a determinant of non-neuronal identity^{3,4}. Here we show that repression of TTK69 protein expression in the neuronal lineage occurs translationally rather than transcriptionally. This translational repression is achieved by a direct interaction between *cis*-acting sequences in the 3' untranslated region of *ttk69* messenger RNA and its *trans*-acting repressor, the RNA-binding protein Musashi (MSI)⁵. Although *msi* can act downstream of Notch, Notch signalling does not affect MSI expression. Thus, Notch signalling is likely to regulate MSI activity rather than its expression. Our results define cell-type-specific translational control of *ttk69* by MSI as a downstream event of Notch signalling in asymmetric cell division.

Mechanosensory bristle development in *Drosophila* is an excellent model system in which to address the molecular mechanisms of asymmetric cell division². Four successive asymmetric cell divisions from a common precursor cell called the sensory organ precursor (SOP) generate a sensory bristle comprising four different non-neuronal support cells and one neuron^{6,7} (Fig. 1a, b and 2f, m). The first asymmetric cell division of SOP into IIa non-neuronal and IIb neuronal precursors is regulated by Notch signalling^{2,8,9}; specific activation of Notch occurs in IIa owing to inhibition in IIb by Numb, an intracellular negative regulator of Notch¹⁰. Activation of Notch signalling results in the appearance of TTK69 protein in the IIa precursor, but not in IIb (refs 3, 10). The expression pattern of TTK69, phenotypes of *ttk69* loss-of-function mutants, and the overexpression of TTK69 (refs 3, 11) suggest that TTK69 is necessary and sufficient to specify the IIa non-neuronal lineage (Fig. 1c, d); however, the mechanism through which Notch signal-

[‡] Present address: Institute of Biological Sciences, University of Tsukuba, Tsukuba 305-8572, Japan.

ling regulates TTK69 expression has remained elusive.

Although TTK69 protein is not detected in the IIb neuronal precursor in which Notch signalling is inactive (Fig. 2a), similar levels of *ttk69* mRNA are present in the IIb precursor and its non-neuronal sibling (Fig. 2b). Using a *lacZ* reporter in a *ttk* enhancer trap line, we confirmed that there are indistinguishable levels of expression between IIa and IIb (Fig. 2c). Selective repression of TTK69 protein expression in the IIb neuronal precursor thus must occur post-transcriptionally, either by post-translational degradation or translational repression.

TTK69 protein also regulates neural development in developing adult photoreceptor cells, where TTK69 expression is controlled by selective degradation in neuronal cells dependent on *seven in absentia* (*sina*) and *phyllopod* functions¹². These two genes are also involved in the cell-fate decision of mechanosensory bristle lineage; loss of either gene function causes a 'double-bristle phenotype', suggesting that the IIb precursor is transformed into IIa (refs 13, 14; and Fig. 2o). However, as the penetrance of the *sina*-null mutant phenotype (12.1%, $n = 107$) is much less than that of the phenotype of overexpression of *ttk69* (ref. 11), other mechanisms must exist for post-transcriptionally regulating *ttk69*.

A good candidate for a repressor of *ttk69* is the *msi* gene product⁵. Loss of *msi* function causes a double-bristle phenotype similar to that of overexpression of TTK69 (36.3%, $n = 327$; Figs 1d and 2g, n). Furthermore, we found that in *msi* mutants TTK69 protein was ectopically expressed in IIb precursors (Fig. 2e), and that a reduction of one copy of the *ttk69* gene dominantly suppressed the *msi* double-bristle phenotype (14.3%, $n = 252$; Fig. 2h). Loss of *ttk* transforms IIa precursor into IIb precursor even in the absence of *msi* (Fig. 2i–l), confirming that *ttk69* acts downstream of *msi* in the first asymmetric cell division.

These results indicate that the *msi* phenotype is caused by a derepression of TTK69 protein expression in the IIb neuronal precursor. This *msi*-dependent pathway acts parallel to the *sina*-dependent pathway, as a *sina msi* double-null mutation shows a much more severe phenotype than either one alone (the production of extra outer support cells in 95.3% of bristles, $n = 86$; Fig. 2p). As *msi* encodes an RNA-binding protein containing two RNA recognition motifs⁵, MSI may act directly on *cis*-acting sequences of the *ttk69* mRNA and regulate expression of TTK69 protein at the translational level.

To identify *cis*-element(s) crucial for the post-transcriptional regulation of *ttk69* mRNA, we took advantage of *ttk*¹, an allele of

ttk with a transposable element (P-element) insertion in the 3' untranslated region (UTR) of *ttk69* mRNA¹⁵ (Fig. 3a). Although this insertion causes the loss of another isoform of TTK (TTK88)¹⁵, it also behaves as a gain-of-function allele of *ttk69*; TTK69 protein was ectopically expressed in the IIb precursor in *ttk*¹ pupal nota (Fig. 3b). In fact, we found that *ttk*¹ display a semi-dominant double-bristle phenotype (5.3% in heterozygous, $n = 245$; 26.2% in homozygous, $n = 233$; Fig. 3c), a phenotype identical to overexpression of either TTK69 (ref. 11) or TTK88 (ref. 3). Thus, the bristle phenotype in *ttk*¹ is caused by ectopic expression of TTK69 rather than the loss of TTK88.

Analysis of the *ttk69* cDNA from *ttk*¹ revealed that its *ttk69* mRNA lacked the distal 80% of the 3' UTR sequence, owing to a read-through into P-element sequences (data not shown). These results indicate the importance of the 3' UTR of *ttk69* mRNA *in vivo* for preventing TTK69 protein expression in the IIb precursor. Thus, the absence of *msi* or the *ttk69* mRNA 3' UTR similarly results in ectopic expression of TTK69 protein in the IIb precursor and the double-bristle phenotype.

To test whether MSI binds *ttk69* mRNA directly, we characterized MSI target sequences *in vitro* and searched for these sequences in the *ttk69* mRNA. We selected random 30- or 50-base-pair RNA polymers by using recombinant glutathione S-transferase (GST)–MSI fusion protein immobilized to glutathione beads, and candidate RNAs were concentrated by five cycles of polymerase chain reaction with reverse transcription (RT–PCR) and *in vitro* transcription¹⁶. We isolated and sequenced 48 independent RNA polymers, and found that all contained a common motif of GU₃–₆(G/AG) (Fig. 3d).

Fifteen such sequences are present in the 3' UTR of the *ttk69*

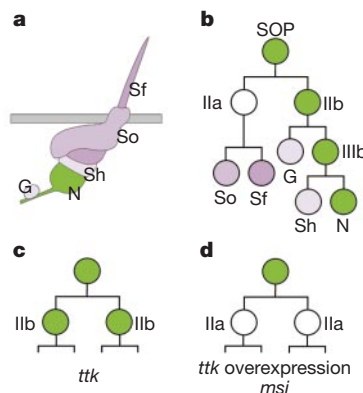


Figure 1 Non-neuronal lineage is determined by TTK69. **a**, Structure of adult mechanosensory bristles. Shown are one neuron (N; green) and four non-neuronal support cells (purple), a shaft (Sf) cell (tricogen), a socket (So) cell (tormogen), a sheath (Sh) cell (thecogen) and glia (G). **b–d**, Cell lineage of mechanosensory bristle in wild-type (**b**), *ttk* loss-of-function mutant (**c**) and overexpression of *ttk69* or *msi* loss-of-function mutants (**d**). Cells with neuronal potential are in green, non-neuronal cells are in purple and white.

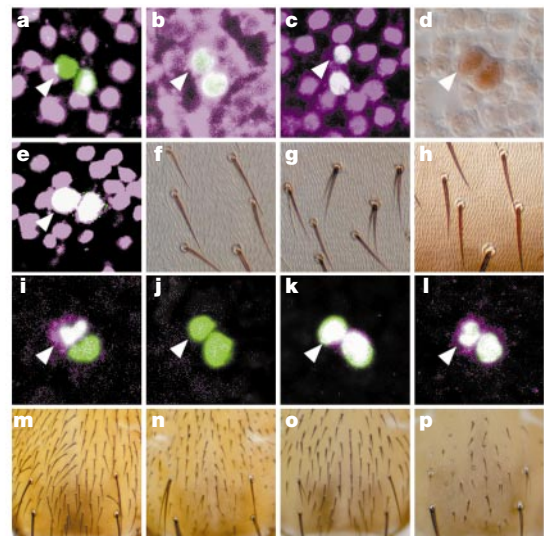


Figure 2 TTK69 is negatively regulated by MSI. **a–e**, Expression pattern of *ttk69* and *msi* at the secondary precursor stage. TTK69 protein (purple, **a**), *ttk69* mRNA (purple, **b**), *lacZ* expression in *ttk*²²¹⁹ enhancer trap line (purple, **c**) and MSI protein (brown, **d**); ectopic TTK69 protein expression (purple) in IIb neuronal precursor in *msi*¹/*msi*¹ (**e**). Cells in the sensory lineage (green) are visualized by nuclear localized β -galactosidase in the *A101* enhancer trap line (**a**, **b**, **e**) or by expression of Cut protein (**c**). Overlap of green and purple is shown in white. Arrowhead indicates the IIb precursor, identified by its relative position to its sister secondary precursor²⁹. **f–h**, Lack of one copy of the *ttk69* gene dominantly suppresses the *msi* phenotype. Mechanosensory bristles in tergite in wild-type (**f**), in *msi*¹/*msi*¹ (**g**) and in *msi*¹/*ttk*²²¹⁹/*msi*¹ (**h**). **i–l**, *ttk* is epistatic to *msi*. Secondary precursor stage of microchaete in wild-type (**i**), *msi*¹/*msi*¹ (**j**), *ttk*²²¹⁹/*ttk*²²¹⁹ (**k**) and *msi*¹/*ttk*²²¹⁹/*msi*¹ (**l**). Cells in the sensory lineage are shown by Cut expression (green). Arrowhead indicates a IIb neuronal precursor expressing its specific marker Prospero protein (purple)⁷. **m–p**, *sina* and *msi* interact synergistically. Notae dissected from wild-type (**m**), *msi*¹/*msi*¹ (**n**), *sina*²/*sina*² (**o**), and *sina*² *msi*¹/*sina*² *msi*¹ (**p**).

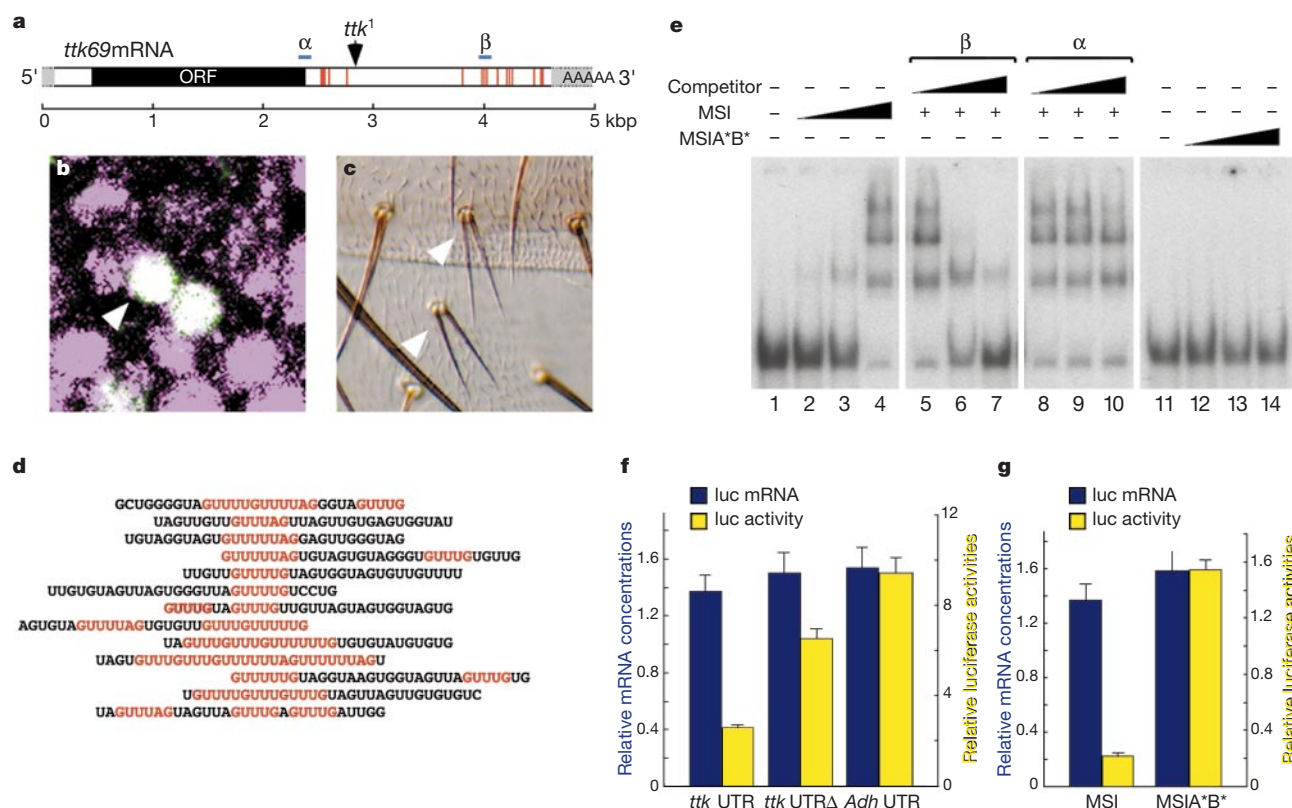


Figure 3 MSI inhibits translation of *ttk69* mRNA by binding to its 3' UTR. **a**, Structure of *ttk69* mRNA deduced from the 4.5-kb cDNA¹⁵. Full-length mRNA is predicted to be 5.0 kb by northern blot analysis¹⁵. The uncharacterized 500-bp sequence (grey), open reading frame (ORF, black), insertion site of the P-element in *ttk*¹ (arrow) and positions of MSI target sequences (red bars) are indicated. **b**, Ectopic TTK69 protein expression (purple) in IIb neuronal precursors in *ttk*¹/*ttk*¹. Both secondary precursors are identified by Cut protein expression (green). Overlap of green and purple is shown in white. Arrowhead indicates the IIb precursor. **c**, *ttk*¹/*ttk*¹ shows a gain-of-function phenotype of *ttk69*: the double-bristle phenotype (arrowhead). **d**, Sequences of 30-bp RNA polymers selected by SELEX. Putative MSI target sequences GU₃₋₆(G/AG) are indicated in red. **e**, MSI binds *ttk69*

mRNA. Each lane contains 10⁴ c.p.m. of ³²P-labelled β-region RNA indicated in **a**. Lanes 1–4, plus 0, 1/27, 1/9 or 1/3 pmol of GST–MSI, respectively; lanes 5–7, plus 1/3 pmol GST–MSI and 4, 20 or 100 pmol non-labelled β-region RNA, respectively; lanes 8–10, plus 1/3 pmol MSI and 4, 20 or 100 pmol α-region RNA, respectively; lanes 11–14, plus 0, 1/27, 1/9 or 1/3 pmol of GST–MSIA*^B*, respectively. **f**, The 3' UTR of *ttk69* mRNA is sufficient to repress translation of the reporter gene in MSI-expressing cultured cells. Relative luciferase activity and relative amount of reporter mRNA are indicated by yellow and blue bars, respectively. **g**, RNA-binding ability of MSI is required for the translational repression.

mRNA (Fig. 3a, red lines), and ten of these are located in the region that is absent in the *ttk69* mRNA produced in the *ttk*¹ mutant. By gel mobility shift assays, we showed that recombinant MSI protein binds specifically to these sequences in *ttk69* mRNA (Fig. 3e). The RNA β-region, containing three MSI target sequences (corresponding to position 3,871–3,880 of *ttk69* cDNA; see Fig. 3a), was bound by MSI but not by a mutant MSI protein with substitutions in the RNA recognition motifs (MSIA*^B)*¹⁷. Binding of the RNA β-region could be competed by a homologous sequence, but not by a region of *ttk69* mRNA lacking MSI-binding motifs (Fig. 3a, 'α'). These results show that MSI is a sequence-specific RNA-binding protein that binds the 3' UTR of *ttk69* mRNA *in vitro*.

To address the functional significance of the binding of MSI to *ttk69* mRNA 3' UTR, we performed a translational reporter assay in *Drosophila* S2 cells (Fig. 3f, g). Luciferase reporter genes that contained the entire 3' UTR of *ttk69* (luc–*ttk*UTR), the proximal 20% of the 3' UTR (luc–*ttk*UTRΔ) or that of alcohol dehydrogenase gene (luc–*Adh*UTR) were introduced in S2 cells that express MSI, and translational and transcriptional outputs were quantified by measuring the enzymatic activities and mRNA levels.

In cells expressing MSI, these three reporter genes produced roughly the same amount of mRNA. When the translational output was measured by luciferase activity, however, the luc–*ttk*UTR reporter produced only 24.7% activity as compared with the *Adh* 3' UTR (Fig. 3f). The luc–*ttk*UTRΔ, which has 5 of the 15 MSI-binding sites present in the *ttk69* 3' UTR and mimics *ttk69* mRNA

produced in *ttk*¹, exhibited only a modest decrease in the translational output (69.8% activity as compared with luc–*Adh*UTR; Fig. 3f). This is consistent with the sharp reduction in MSI-dependent translational repression in the *ttk*¹ mutant, as the penetrance of *ttk*¹ homozygous is less than that of *msi* null homozygous. The effect of *ttk69* 3' UTR was dependent on the RNA-binding activity of MSI; in cells expressing MSI lacking RNA-binding activity (MSIA*^B)*, translational repression of the *ttk69* 3' UTR was not observed (Fig. 3g). The 3' UTR of *ttk69* mRNA therefore confers MSI-dependent translational repression in this assay system. Together, these results indicate that MSI inhibits translation of *ttk69* mRNA in IIb precursors by binding to its 3' UTR.

Although the translational inhibitory effect of MSI on *ttk69* mRNA is specific to the IIb precursors, MSI protein was present in both IIa and IIb precursors (Fig. 2d). Thus, IIa precursors must somehow be able to escape the action of MSI as a translational repressor of TTK69, probably through the effect of Notch signalling. Loss of Notch function in the SOP lineage causes the transformation of IIa precursor into IIb, with mutants showing a balding phenotype owing to loss of socket and shaft cells⁸. This phenotype in IIa precursor is dependent on MSI activity; loss of both *Notch* and *msi* function results in a dense double-bristle phenotype (Fig. 4a–c) with no neurons in the subepidermal layer (data not shown), indicating that the IIb precursor took the non-neuronal fate. *msi* is thus epistatic to *Notch* during the asymmetric cell division giving

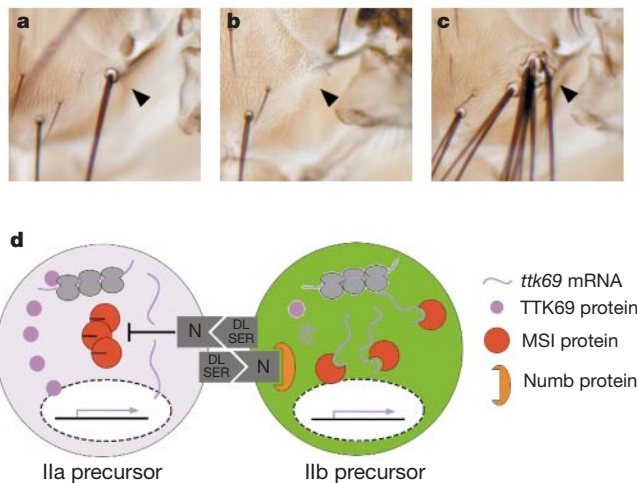


Figure 4 MSI acts downstream to Notch. **a–c**, *msi* is epistatic to *Notch*. A mechanosensory bristle in lateral side of notum in wild-type (**a**), *Gal4¹⁰⁹/UAS-Notch^PN* (**b**) and *Gal4¹⁰⁹/UAS-Notch^PN*; *msi¹/msi¹* (**c**). **d**, Model of asymmetric cell division based on *ttk69* post-transcriptional regulation. SINA-dependent degradation also contributes to Ila-specific expression of TTK69.

rise to Ila and I Ib precursors. Taken together, we propose that Notch inhibits MSI activity in the Ila precursor (non-neuronal cell), thus allowing translation of *ttk69* mRNA, whereas in the I Ib precursor (neuronal cell), where Notch is inactivated by Numb, MSI prevents *ttk69* translation (Fig. 4d).

mRNA localization and translational control is a major strategy for establishing embryonic pattern in early development, when the absence of zygotic gene transcription necessitates gene regulation at the post-transcriptional level¹⁸. Once zygotic gene expression starts, transcriptional gene regulation has been thought to be a principal mechanism to make differential gene expression in subsequent asymmetric cell divisions. Our results show that cell-type-specific translational control of gene expression acts as a downstream event of Notch signalling during post-embryonic development and can direct cell fates in a small subset of cells in the nervous system. Homologues of Notch and MSI are highly expressed in the neural progenitor cells of the vertebrate nervous system^{19–21}, which suggests the evolutionary conservation of this translational gene regulation in the developing nervous system. □

Methods

Immunohistochemistry and *in situ* hybridization

Antibody staining and *in situ* hybridization were done as described²². We used the following antibodies: rat anti-TTK69 (ref. 4); rat anti-MSI monoclonal 3A5 (ref. 17); mouse anti-Prospero monoclonal MR1a (ref. 23); rabbit anti-Cut (ref. 24); mouse anti-Cut monoclonal 2B10 and mouse anti- β -galactosidase monoclonal 40-1a (Developmental Studies Hybridoma Bank, Univ. Iowa); rabbit anti- β -galactosidase (Cappel). *ttk69* cDNA¹⁵ was digested by *Bam*HI to prepare a template DNA for *in vitro* transcription of the *ttk69*-specific antisense RNA probe. Immunofluorescence was visualized with a BioRad MRC 1024 confocal microscope.

Genetics

Oregon-R was used as the wild-type and enhancer trap line *A101* (ref. 25) for identification of SOP and its progeny. We also used the lines: *msi¹* (ref. 5), *ttksm* (ref. 3), *ttk¹* (ref. 15), *ttk²¹⁹* (ref. 12), *sina²* (ref. 13), *sina³* (ref. 13), *GAL4¹⁰⁹⁻⁶⁸* (ref. 9) and *UAS-Notch^{ECN}* (ref. 26). *ttk* or *msi* *ttk* double-mutant clones were induced by heat shocking larvae of genotype *y w hsFLP;FRT82B ttksm or FRT82B msi¹ ttksm/FRT82B ubGFP81F M(3)95A²*. FLP-induced recombination produces clones of cells in pupal notum that lack expression of green fluorescent protein.

RNA selection

RNA selection was done as described¹⁶ with minor modifications. Recombinant GST–MSI fusion protein was produced from pGEX-2T vector containing residues 129–356 of MSI¹. For the random DNA library, oligonucleotides containing a 30- or 50-bp random sequence and primer-binding sites at both ends (5'-GGGAAGATCTCGACCAAG-

N30-TATGTGCGTCTACATGGATCCTCA-3') were synthesized and amplified by PCR, with a forward primer containing a T7 promoter sequence. We prepared a random RNA pool by *in vitro* transcription using the random DNA library and T7 RNA polymerase. Glutathione–Sephacrose 4B resins (50 ml; 50% slurry) conjugated with 5 mg of purified GST–MSI fusion protein were incubated with 10 mg of pooled RNAs for 20 min at room temperature in 0.5 M LiCl, 20 mM Tris–HCl pH 7.5, 1 mM MgCl₂. After five washes, bound RNAs were released by phenol extraction, reverse transcribed with M-MLV reverse transcriptase (GibcoBRL), and amplified by PCR using forward and reverse primers. We repeated this process four times, and subcloned and sequenced the amplified DNAs.

Gel mobility shift assay

Gel mobility shift assays were performed with indicated amounts of GST–MSI or GST–MSIA**B** fusion protein in 16 ml of binding buffer (20 mM KCl, 150 mM NaCl, 50 mM Tris–HCl pH 7.5, 2 mM EGTA, 0.05% NP-40, 20 units RNasin (Promega), 6 mM dithiothreitol). Roughly 5 fmol (10⁴ c.p.m.) of ³²P-labelled RNAs or non-labelled competitor RNA were produced by *in vitro* transcription using subcloned *ttk69* cDNAs. We incubated the RNAs with GST–fusion proteins for 30 min at room temperature in the binding buffer. After incubation, the mixtures were loaded onto a 6% polyacrylamide gel (0.5× TBE buffer). The gel was dried and exposed to XAR autoradiography film (Kodak).

Translation assay

Culture of *Drosophila* S2 cells and transfection of plasmid DNA were done as described²⁷. The *Photinus* luciferase GL-3 (Promega) coding sequence was cloned into pRmHa-3 *Metallothionein* vector²⁸ with the 3' UTR of *Adh* to make the luc–*Adh*UTR reporter plasmid; GL-3 coding sequence followed by *ttk69* 3' UTR (2230–4524 or 2230–2741)¹⁵ was cloned into pRmHa-3 vector for the respective luc–*ttk*UTR or luc–*ttk*UTRΔ reporter plasmid. The internal control *Metallothionein* reporter was also constructed from *Remilla* luciferase RL and *Adh* 3' UTR. The MSI or MSIA**B** (ref. 17) coding sequence was cloned into pAct actin promoter vector for the respective MSI or MSIA**B** expression vectors. After co-transfection cells were allowed to recover and express MSI or MSIA**B** for 48 h, then expression of the *Metallothionein* reporter construct was induced by adding CuSO₄ to 0.7 mM for 6 h. The mRNA levels and activities of both *Photinus* and *Remilla* luciferases were quantified by a northern ELISA system (Rosh Diagnostics) and the Dual-Luciferase Reporter Assay system (Promega), respectively. We carried out transfections at least five times, and values are reported as means $\pm$ s.d..

Received 5 September 2000; accepted 8 February 2001.

- Horvitz, H. R. & Herskowitz, I. Mechanisms of asymmetric cell division: two Bs or not two Bs, that is the question. *Cell* **68**, 237–255 (1992).
- Jan, Y. N. & Jan, L. Y. Asymmetric cell division. *Nature* **392**, 775–778 (1998).
- Guo, M., Bier, E., Jan, L. Y. & Jan, Y. N. Tramtrack acts downstream of numb to specify distinct daughter cell fates during asymmetric cell divisions in the *Drosophila* PNS. *Neuron* **14**, 913–925 (1995).
- Giesen, K. *et al.* Glial development in the *Drosophila* CNS requires concomitant activation of glial and repression of neuronal differentiation genes. *Development* **124**, 2307–2316 (1997).
- Nakamura, M., Okano, H., Blendy, J. A. & Montell, C. Musashi, a neural RNA-binding protein required for *Drosophila* adult external sensory organ development. *Neuron* **13**, 67–81 (1994).
- Gho, M., Bellaiche, Y. & Francois, S. Revisiting the *Drosophila* microchaete lineage: a novel intrinsically asymmetric cell division generates a glial cell. *Development* **126**, 3573–3584 (1999).
- Reddy, G. V. & Rodrigues, V. A glial cell arises from an additional division within the mechanosensory lineage during development of the microchaete on the *Drosophila* notum. *Development* **126**, 4617–4622 (1999).
- Hartenstein, V. & Posakony, J. W. A dual function of the Notch gene in *Drosophila* sensillum development. *Dev. Biol.* **142**, 13–30 (1990).
- Zeng, C., Younger-Shepherd, S., Jan, L. Y. & Jan, Y. N. Delta and Serrate are redundant Notch ligands required for asymmetric cell divisions within the *Drosophila* sensory organ lineage. *Genes Dev.* **12**, 1086–1091 (1998).
- Guo, M., Jan, L. Y. & Jan, Y. N. Control of daughter cell fates during asymmetric division: interaction of Numb and Notch. *Neuron* **17**, 27–41 (1996).
- Ramaekers, G. *et al.* Lineage and fate in *Drosophila*: Role of the gene tramtrack in sense organ development. *Dev. Genes Evol.* **207**, 97–106 (1997).
- Li, S., Li, Y., Carthew, R. W. & Lai, Z. C. Photoreceptor cell differentiation requires regulated proteolysis of the transcriptional repressor tramtrack. *Cell* **90**, 469–478 (1997).
- Carthew, R. W. & Rubin, G. M. seven in absentia, a gene required for specification of R7 cell fate in the *Drosophila* eye. *Cell* **63**, 561–577 (1990).
- Chang, H. C. Y. *et al.* phyllopod functions in the fate determination of a subset of photoreceptors in *Drosophila*. *Cell* **80**, 463–472 (1995).
- Xiong, W. C. & Montell, C. tramtrack is a transcriptional repressor required for cell fate determination in the *Drosophila* eye. *Genes Dev.* **7**, 1085–1096 (1993).
- Buckanovich, R. J. & Darnell, R. B. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol. Cell. Biol.* **17**, 3194–3201 (1997).
- Hirota, Y. *et al.* Musashi and Seven in absentia downregulate Tramtrack through distinct mechanisms in *Drosophila* eye development. *Mech. Dev.* **87**, 93–101 (1999).
- Gray, N. K. & Wickens, M. Control of translation initiation in animals. *Annu. Rev. Cell. Dev. Biol.* **14**, 399–458 (1998).
- Sakakibara, S. *et al.* mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell. *Dev. Biol.* **176**, 230–242 (1996).
- Johansson, C. B. *et al.* Identification of a neural stem cell in the adult mammalian central nervous. *Cell* **96**, 25–34 (1999).
- Gaiano, N., Nye, J. S. & Fishell, G. Radial glial identity is promoted by Notch1 signaling in the murine forebrain. *Neuron* **26**, 395–404 (2000).
- Goto, S. & Hayashi, S. Cell migration within the embryonic limb primordium of *Drosophila* as revealed by a novel fluorescence method to visualize mRNA and protein. *Dev. Genes Evol.* **207**, 194–198 (1997).

23. Spana, E. P. & Doe, C. Q. The prospero transcription factor is asymmetrically localized to the cell cortex during neuroblast mitosis in *Drosophila*. *Development* **121**, 3187–3195 (1995).
24. Blochliger, K., Bodmer, R., Jan, L. Y. & Jan, Y. N. Patterns of expression of cut, a protein required for external sensory organ development in wild-type and cut mutant *Drosophila* embryos. *Genes Dev.* **4**, 1322–1331 (1990).
25. Ghysen, A. & O'Kane, C. Neural enhancer-like elements as specific cell markers in *Drosophila*. *Development* **105**, 35–52 (1989).
26. Jacobsen, T. L., Brennan, K., Arias, A. M. & Muskavitch, M. A. Cis-interactions between Delta and Notch modulate neurogenic signalling in *Drosophila*. *Development* **125**, 4531–4540 (1998).
27. Jin, M. H., Sawamoto, K., Ito, M. & Okano, H. The interaction between the *Drosophila* secreted protein Argos and the epidermal growth factor receptor inhibits dimerization of the receptor and binding of secreted spitz to the receptor. *Mol. Cell. Biol.* **20**, 2098–2107 (2000).
28. Bunch, T. A., Grinblatt, Y. & Goldstein, L. S. Characterization and use of the *Drosophila* metallothionein promoter in cultured *Drosophila melanogaster* cells. *Nucleic Acids Res.* **16**, 1043–1061 (1988).
29. Gho, M., Lecourtis, M., Geraud, G., Posakony, J. W. & Schweisguth, F. Subcellular localization of Suppressor of Hairless in *Drosophila* sense organ cells during Notch signalling. *Development* **122**, 1673–1682 (1996).

Acknowledgements

We thank M. Nakamura, Y. N. Jan, C. Montell, S. Hayashi, A. Travers, C. Q. Doe and the Developmental Studies Hybridoma Bank, University of Iowa, for fly strains and/or DNA clones and antibodies for fly strains. We are grateful to K. Tei and D. Yamamoto for advice on the translation assay. We also thank M. Nakamura, S. Goto, M. Jindra and all members of the Okano, Hiromi, Hotta and Hayashi laboratories for helpful discussions. This work was supported by grants from the ministry of Education, Science, Sports, and Culture of Japan, and Japan Science and Technology Corporation.

Correspondence and requests for materials should be addressed to M.O. (e-mail: maokabe@lab.nig.ac.jp). or H.O. (e-mail: hidokano@sc.itc.keio.ac.jp)

Exploitation of syndecan-1 shedding by *Pseudomonas aeruginosa* enhances virulence

Pyong Woo Park^{*†}, Gerald B. Pier[‡], Michael T. Hinkes^{*} & Merton Bernfield^{*}

^{*} Division of Newborn Medicine, Department of Pediatrics, Children's Hospital and [‡] Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

Cell-surface heparan sulphate proteoglycans (HSPGs) are ubiquitous and abundant receptors/co-receptors of extracellular ligands^{1,2}, including many microbes^{3–10}. Their role in microbial infections is poorly defined, however, because no cell-surface HSPG has been clearly connected to the pathogenesis of a particular microbe. We have previously shown that *Pseudomonas aeruginosa*, through its virulence factor LasA, enhances the *in vitro* shedding of syndecan-1—the predominant cell-surface HSPG of epithelia¹¹. Here we show that shedding of syndecan-1 is also activated by *P. aeruginosa* *in vivo*, and that the resulting syndecan-1 ectodomains enhance bacterial virulence in newborn mice. Newborn mice deficient in syndecan-1 resist *P. aeruginosa* lung infection but become susceptible when given purified syndecan-1 ectodomains or heparin, but not when given ectodomain core protein, indicating that the ectodomain's heparan sulphate chains are the effectors. In wild-type newborn mice, inhibition of syndecan-1 shedding or inactivation of the shed ectodomain's heparan sulphate chains prevents lung infection. Our findings uncover a pathogenetic mechanism in which

a host response to tissue injury—syndecan-1 shedding—is exploited to enhance microbial virulence apparently by modulating host defences.

To determine whether cell-surface HSPGs are important host determinants for microbial pathogenesis, we initially examined the virulence of *P. aeruginosa* in an intranasally induced lung infection model¹² using 7-day-old syndecan-1 null (*synd-1^{-/-}*) and wild-type control (*synd-1^{+/+}*) mice. We used this model because naive newborn mice are susceptible to *P. aeruginosa* lung infection through intranasal administration¹³, whereas adult mice require anaesthetics to establish *P. aeruginosa* lung infection through this route¹⁴. The role of syndecan-1 in binding and regulating the activity of various ligands (such as matrix components, growth factors, cytokines and morphogens) is well established². The *synd-1^{-/-}* mice are phenotypically identical to wild-type animals under normal laboratory conditions^{1,15}, and thus are an ideal model to assess the role of syndecan-1 in microbial pathogenesis. *P. aeruginosa*, a clinically important Gram-negative bacterium, was chosen as the test pathogen because, like many other microbial pathogens, it targets the host epithelia early in infection, where the principal cell-surface HSPG is syndecan-1. Furthermore, we previously found that *P. aeruginosa* enhances the shedding of syndecan-1 ectodomains from cultured epithelial cells¹¹.

Seven-day-old *synd-1^{-/-}* mice markedly resisted infection by intranasally inoculated *P. aeruginosa* (strain PAO1) (Fig. 1a). With the largest inoculum of *P. aeruginosa* given to BALB/C mice (Fig. 1a), the lungs and spleens of wild-type *synd-1^{+/+}* mice were extensively colonized by *P. aeruginosa*, reflecting pneumonia and bacteraemia, and 10 of the 27 (37%) of the wild-type *synd-1^{+/+}* animals compared with none of the 13 *synd-1^{-/-}* animals died from the infection ($P = 0.01$, Fisher's exact test). In addition, colonization of the lungs and spleens of *synd-1^{-/-}* mice by *P. aeruginosa* was reduced by roughly 10^4 and 10^3 c.f.u., respectively, compared with that of *synd-1^{+/+}* mice. Consistent with these findings, histological analyses at 8 h and 20 h showed florid pneumonia with innumerable *P. aeruginosa* bacilli in the lungs of wild-type mice, whereas the lungs of *synd-1^{-/-}* mice showed few or no inflammatory foci (Fig. 1b). Similar results were obtained with *synd-1^{+/+}* and *synd-1^{-/-}* mice in the C57BL/6 background (Fig. 1a), and also by using the *P. aeruginosa* clinical isolates 6077 (ref. 14) and BL2 (ref. 11). Notably, *synd-1^{-/-}* mice were as susceptible as wild-type *synd-1^{+/+}* mice to infection after intraperitoneal inoculation (see Supplementary Information). Together, these findings show that syndecan-1 is an important host determinant in *P. aeruginosa* infection of newborn mouse lungs and possibly in other infections in which the bacteria first encounter the host's epithelia—a cell type whose predominant HSPG is syndecan-1.

Because many pathogens exploit epithelial cell-surface HSPGs for their attachment^{3,5–10}, we investigated the possibility that syndecan-1 is an attachment site for *P. aeruginosa*, and that *synd-1^{-/-}* mice resist *P. aeruginosa* infection because they lack the syndecan-1 attachment sites. Consistent with previous findings¹⁶, however, we observed that *P. aeruginosa* does not bind to syndecan-1 labelled by either ¹²⁵I (protein labelling) or ³⁵S (heparan sulphate (HS) labelling), and that adhesion of *P. aeruginosa* to cultured lung epithelial cells is not affected by excess heparin, which is a functional pharmaceutical mimic of the highly sulphated domains of HS (see Supplementary Information). These results suggest that *P. aeruginosa* does not exploit syndecan-1 as an attachment site.

Shedding of syndecan-1 ectodomain is activated as one of the host responses to tissue injury¹. Although shed syndecan-1 ectodomains are not detected in biological fluids of healthy patients, they are easily detected in tissue injury fluids such as tracheal aspirate fluids of ventilated pre-term infants and skin wound fluids^{17,18}. Thus, we speculated that *synd-1^{-/-}* mice resist infection because they lack shed syndecan-1 ectodomains. If so, administering purified shed syndecan-1 ectodomains to the *synd-1^{-/-}* mice might reverse

[†] Present address: Section of Infectious Diseases, Baylor College of Medicine, One Baylor Plaza, BCM 286, Room N1319, Houston, Texas 77030, USA.

their resistance and make them more susceptible to infection.

To test this hypothesis, 7-day-old synd-1^{-/-} mice were intranasally infected with *P. aeruginosa* in the absence or presence of purified shed syndecan-1 ectodomain, heparin or control reagents. synd-1^{-/-} mice given shed syndecan-1 ectodomain or heparin, but not HS-free syndecan-1 core protein or chondroitin sulphate, became susceptible to *P. aeruginosa* infection (Fig. 2a). Six out of ten animals died in the cohort administered shed ectodomain or heparin, whereas none of the animals died in groups co-inoculated with HS-free syndecan-1 core protein (0/7) or chondroitin sulphate (0/7) ($P = 0.017$, Fisher's exact test). In the absence of bacteria, inoculation of these reagents did not cause any apparent ill effects.

Consistent with our findings that shed syndecan-1 does not interact with *P. aeruginosa*, pre-incubation of the bacteria with heparin (5 ng μl^{-1}) for 30 min did not enhance virulence if the heparin was removed by washing before inoculation (see Supplementary Information). Furthermore, the enhanced susceptibility of synd-1^{-/-} mice to infection after administration of the shed ectodomain or heparin was reversed when these mice were co-inoculated with protamine (200 ng per 7 μl)—an agent that binds and neutralizes heparin and HS (Fig. 2b). Although protamine has antibacterial activities at high concentrations, growth and viability of *P. aeruginosa* were not affected by protamine, nor by any of the other exogenously added agents at the concentrations used in our studies.

Administering purified shed syndecan-1 ectodomains and heparin also enhanced the virulence of *P. aeruginosa* in the wild-type synd-1^{+/+} background by increasing mortality values from 37% (10/27) to 70% (7/10, $P = 0.06$, Fisher's exact test) and 88% (15/17, $P = 0.0008$, Fisher's exact test), respectively. These results indicate that synd-1^{-/-} mice resist *P. aeruginosa* lung infection because their lungs lack the shed syndecan-1 ectodomains, that syndecan-1 ectodomains enhance bacterial virulence in the wild-type synd-1^{+/+} background, and that the active moieties are the syndecan-1 HS chains.

If activation of syndecan-1 ectodomain shedding is indeed important for *P. aeruginosa* lung pathogenesis, then *P. aeruginosa*

may possess the ability to stimulate syndecan-1 shedding *in vivo* to enhance its virulence. Using cultured epithelial cells, we previously found that one of the *P. aeruginosa* factors that enhances syndecan-1 shedding is LasA¹¹—a known virulence factor for *P. aeruginosa* lung infection^{19,20}. Syndecan-1 shedding can be inhibited by peptide hydroxamate inhibitors (such as BB1101) of the endogenous metalloproteinase responsible for cleaving cell-surface syndecan-1 (ref. 11). However, these inhibitors can also act on other zinc metalloproteinases of the MMP (matrix metalloproteinase) and ADAM (a disintegrin and metalloproteinase) families, and potentially zinc metalloproteinases of microbial pathogens.

To determine whether *P. aeruginosa* stimulates syndecan-1 shedding *in vivo*, we inoculated 7-day-old wild-type synd-1^{+/+} mice with control media, *P. aeruginosa* or purified LasA, in the absence or presence of BB1101, and measured the relative amount of soluble, shed syndecan-1 and -4 ectodomains in the lung. Syndecan-4 is another cell-surface HSPG expressed by lung epithelia, albeit at a much lower amount. The relative amount of shed syndecan-1 ectodomains increased significantly in response to both *P. aeruginosa* and LasA, and this stimulation was inhibited by BB1101 (Fig. 3). In contrast, the amount of shed syndecan-4 ectodomain was not affected by either *P. aeruginosa* or LasA inoculation relative to the media control, suggesting that *P. aeruginosa* specifically targets syndecan-1 shedding *in vivo*.

Our data uncover a pathogenetic mechanism of *P. aeruginosa* in which the bacteria specifically stimulate syndecan-1 shedding to generate shed syndecan-1 ectodomains that, through their HS chains, can enhance bacterial virulence. If this mechanism is indeed important for *P. aeruginosa* lung pathogenesis, then inhibition of syndecan-1 shedding or the activity of the shed ectodomain's HS chains should interfere with *P. aeruginosa* infection. To test this hypothesis, we examined the effects of BB1101, protamine and an HS-degradative enzyme (heparitinase II) on *P. aeruginosa* virulence in the intranasal lung infection model using 7-day-old wild-type syndecan-1^{+/+} mice.

Co-inoculation with either BB1101 (20 μM) or protamine (200 ng per 7 μl) reduced *P. aeruginosa* colonization of the lungs

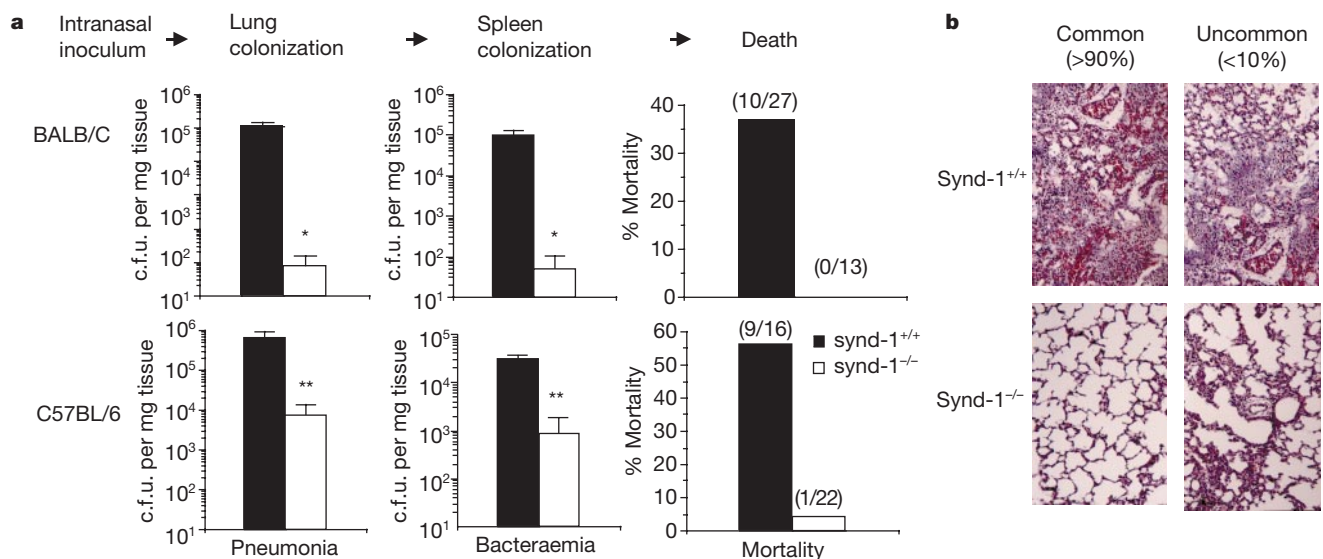


Figure 1 Synd-1^{-/-} mice resist intranasally induced *P. aeruginosa* infection. **a**, Seven-day-old synd-1^{-/-} or synd-1^{+/+} mice in either a BALB/C or C57BL/6 genetic background were inoculated intranasally with *P. aeruginosa* strain PAO1 ((1.8–2.1) × 10⁹ c.f.u. per 7 μl). This relatively high inoculum of *P. aeruginosa* is needed to establish infection in wild-type synd-1^{+/+} animals²⁸. Infection was assessed after 20 h by quantifying colonized bacteria in the lung and spleen, and by observing mortality. Results for lung and spleen colonization represent mean ± s.e. (* $P \leq 0.04$, ** $P \leq 0.03$, Student's *t*-test).

Comparison of mortality percentages: mice in the BALB/C background, $P = 0.010$; mice in the C57BL/6 background, $P = 0.0005$ (Fisher's exact test). **b**, Upper left lobes of infected synd-1^{+/+} and synd-1^{-/-} lungs in the BALB/C background were isolated after 8 h, fixed in 4% paraformaldehyde/PBS for 48 h, embedded in paraffin, sectioned in 5- μm slices, and stained with haematoxylin–eosin. Two regions from each group, representing common (> 90% of specimen) and uncommon (< 10% of specimen) areas, are shown.

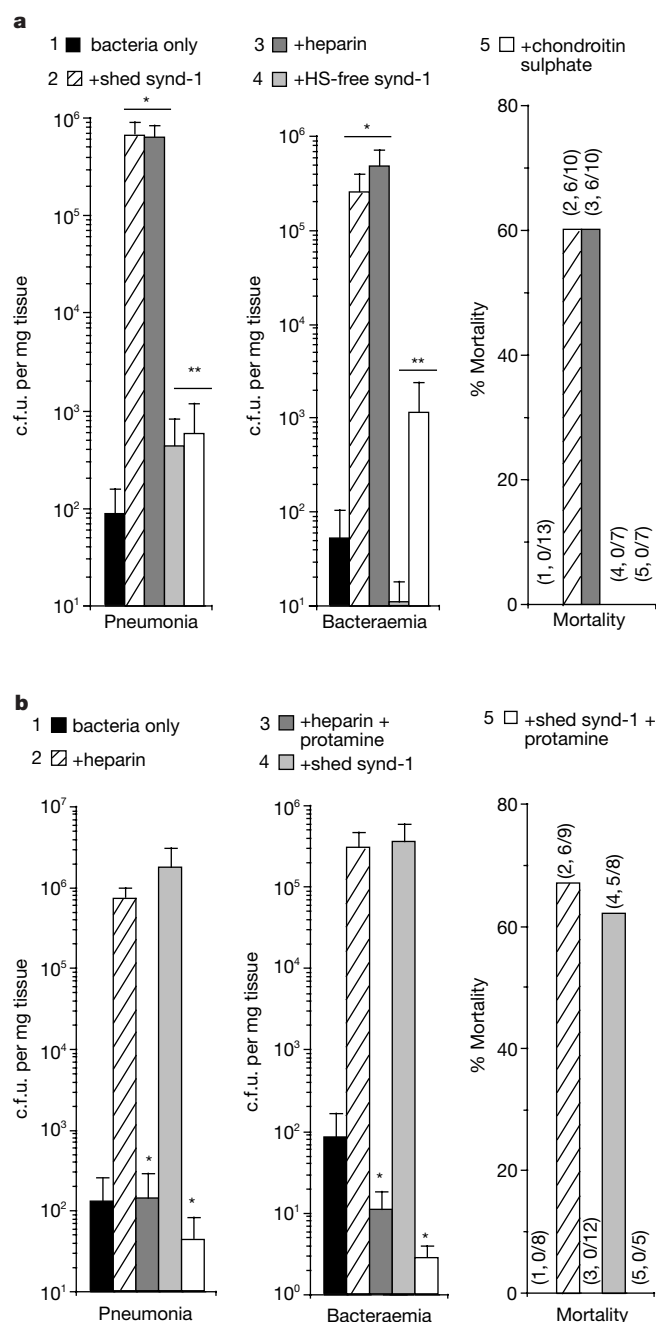


Figure 2 Resistance of the synd-1^{-/-} mice to *P. aeruginosa* infection can be reversed by administering purified shed syndecan-1 ectodomains or heparin. **a**, Seven-day-old synd-1^{-/-} mice were inoculated with $(1.8-2.2) \times 10^9$ c.f.u. per 7 μ l of *P. aeruginosa* PAO1 without or with purified shed syndecan-1 ectodomain (80 ng), heparin (35 ng), HS-free syndecan-1 core protein (80 ng), or chondroitin sulphate C (35 ng), and assessed for colonization as in Fig. 1. Reagents and bacteria were mixed immediately before inoculation and were incubated for less than 10 min. Data for pneumonia and bacteraemia represent mean $\pm$ s.e. (* $P \leq 0.03$, ** $P > 0.20$ versus mice given only bacteria; Student's *t*-test). Comparison of mortality percentages: synd-1^{-/-} mice plus shed syndecan-1 or plus heparin versus control mice, $P = 0.002$ (Fisher's exact test). **b**, Seven-day-old synd-1^{-/-} mice were inoculated with *P. aeruginosa* PAO1 in 7 μ l TSB without or with heparin (35 ng), heparin (35 ng) plus protamine (200 ng), purified shed syndecan-1 ectodomain (80 ng), or purified ectodomain (80 ng) plus protamine (200 ng). Results for pneumonia and bacteraemia represent mean $\pm$ s.e. (* $P \leq 0.015$, mice given heparin versus heparin plus protamine, and mice given shed syndecan-1 versus shed syndecan-1 plus protamine; Student's *t*-test). Comparison of mortality percentages: synd-1^{-/-} mice given heparin versus mice given heparin plus protamine, $P = 0.002$; synd-1^{-/-} mice given shed syndecan-1 versus mice given shed syndecan-1 plus protamine, $P = 0.04$ (Fisher's exact test).

and spleens, and prevented death (Fig. 4). Pre-treatment of the animals with heparitinase II (0.3 mU per 7 μ l), but not with the chondroitin sulphate-digesting enzyme (chondroitinase ABC, 0.3 mU per 7 μ l), also significantly reduced lung virulence of *P. aeruginosa*. None of the reagents at the concentration tested affected bacterial growth or viability. These results correlate syndecan-1 shedding to virulence, establish shed HS chains as important host-derived effectors that can enhance *P. aeruginosa* virulence in the lung, and suggest possible therapeutic approaches.

Our findings show that *P. aeruginosa* exploits syndecan-1 to enhance its virulence as a shed ectodomain, and not as a cell-surface HSPG. Exactly how HS chains of syndecan-1 ectodomains enhance *P. aeruginosa* virulence remains to be delineated, but we have found that the ectodomains do not interact directly with *P. aeruginosa*, and that pre-incubation of bacteria with heparin does not enhance virulence if heparin is removed before inoculating the animals. These results indicate that the effect of shed ectodomains is not on the bacteria, and suggest that shed ectodomains enhance *P. aeruginosa* virulence by interfering with host defence.

Shed syndecan-1 ectodomains may interfere with host defences by interacting with different agents of the innate defence system. Purified syndecan-1 ectodomains, through their HS chains, bind tightly to cationic antimicrobial peptides of the (Pro/Arg)-rich cathelicidin family (such as batenecins 5 and 7, or PR-39) and inhibit their antibacterial activities (see Supplementary Information). Because several other antimicrobials in the lung are highly cationic (such as lysozyme, lactoferrin, β -defensins, LL-37)^{21,22}, shed ectodomains may enhance virulence by inhibiting the activity of these agents. Shed ectodomains also bind to neutrophil elastase and cathepsin G (ref. 23). Elastase, at least, has been shown to be important in defending the host against Gram-negative bacterial sepsis²⁴. Furthermore, we have preliminary evidence showing that syndecan-1 ectodomains bind to surfactant proteins A and D in a calcium-dependent manner. These surfactants belong to the collectin family of host defence molecules, and are critical in protecting the host from microbial lung infections, especially *P. aeruginosa*^{25,26}. Binding of shed ectodomains would be expected to prevent collectin aggregation and thus reduce their host defence activity. Finally, soluble HS can inhibit the activity of several cytokines involved in phagocyte recruitment²⁷, implying that HS chains of shed ectodomains may act similarly. Together, these findings suggest that inhibition of soluble, innate host defence factors may be one of

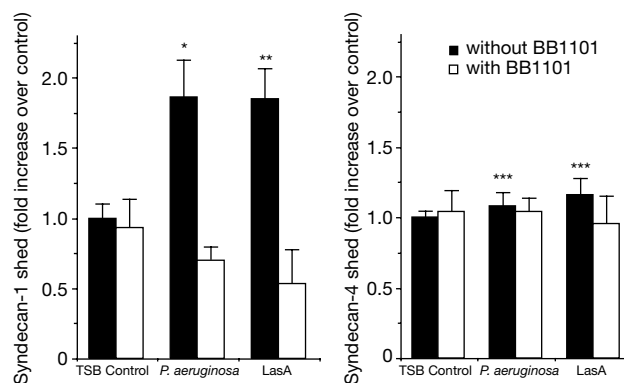


Figure 3 *P. aeruginosa* specifically increases the amount of shed syndecan-1 ectodomains in the lung. Wild-type synd-1^{+/+} mice (7-day-old) were inoculated with either 7 μ l TSB ($n = 6$), 7 μ l TSB containing 1.9×10^9 c.f.u. of *P. aeruginosa* PAO1 ($n = 6$), or 7 μ l TSB containing 500 ng of purified LasA ($n = 6$) in the absence or presence of 20 μ M BB1101 ($n = 3$ for each group). After 8 h, the lungs from each animal were isolated and the soluble syndecan-1 (left) and -4 (right) ectodomains determined as described in Methods. Results represent mean $\pm$ s.e. (* $P = 0.03$, ** $P = 0.01$, *** $P > 0.15$; Student's *t*-test).

the underlying mechanisms by which shed syndecan-1 shifts host defence to favour *P. aeruginosa* pathogenesis.

Another important issue is whether this mechanism is used by other microbes for their pathogenesis. Our results show that the specificity of this mechanism is maintained at the level of activating syndecan-1 shedding to generate shed ectodomains. In light of this key event, we previously found that another major lung bacterial pathogen, *Staphylococcus aureus*, also potentially stimulates syndecan-1 shedding¹¹, and preliminary results show that the synd-1^{-/-} mice also resist intranasally induced *S. aureus* lung infection; however, we do not yet know whether the synd-1^{-/-} mice resist *S. aureus* infection because of the lack of shed ectodomains. Our findings also suggest that clinical circumstances that result in enhanced shedding of syndecan-1 ectodomains may predispose patients to bacterial infections. For example, both tracheal aspirates of ventilated pre-term infants and skin wound fluids contain abnormally high levels of shed syndecan-1 ectodomains^{17,18}, and both *P. aeruginosa* and *S. aureus* are major pathogens in ventilated lungs and skin wounds. Finally, as both *P. aeruginosa* and *S. aureus* represent chief lung pathogens for cystic fibrosis patients, it will be of interest to determine whether the lungs of these patients have increased levels of shed syndecan-1 ectodomains.

We have identified syndecan-1 as a specific cell-surface HSPG that has a pivotal role in *P. aeruginosa* lung pathogenesis. Our results extend the biological function of syndecan-1 from that of a cell-surface molecule regulating cellular activities, such as adhesion, proliferation and differentiation, to that of a soluble HSPG capable of modulating microbial pathogenesis and host defence.

Note added in proof: After this paper was accepted, a publication

appeared²⁹ demonstrating that another class of solubilized glycosaminoglycan, dermatan sulphate, can bind and inhibit the bactericidal activity of α -defensin, which is similar to the binding and inhibition of cathelicidin antimicrobial peptides by shed syndecan-1 ectodomains reported here. □

Methods

Animals

Synd-1^{-/-} mice were generated by homologous recombination on a 129/SvJ genetic background¹⁵. We bred chimaeric mice with C57BL/6 mates until the null mutation was homozygous. Null mice in the C57BL/6 and BALB/C genetic backgrounds were back-crossed four and six times to their respective backgrounds, and were maintained by interbreeding at our facility. All of the experiments using synd-1^{-/-} mice in this study, except for the C57BL/6 experiments of Fig. 1a, were performed with N6 BALB/C synd-1^{-/-} mice. Wild-type synd-1^{+/+} mice were either obtained from The Jackson Laboratory (Bar Harbor, Maine, USA), or interbred and maintained in our laboratory. We housed and studied animals under protocols approved by the Institutional Animal Care and Use Committee in the animal facility of Children's Hospital (Boston, MA).

Lung infection assay

A described lung infection model¹² was used to infect 7-day-old newborn mice. Seven-day-old mice are susceptible to *P. aeruginosa* lung infection without the use of anaesthetics¹³, whereas adult mice require anaesthetics to establish *P. aeruginosa* lung infection through intranasal administration¹⁴; thus, newborn mice were used in all experiments to avoid the unpredictable effects of anaesthesia on syndecan-1 shedding. Our results have not yet been extended to adult animals. In brief, *P. aeruginosa* was grown overnight in TSB, and the bacterial concentration was approximated by measuring absorbance at 600 nm. After washing, the concentration was adjusted to roughly 2×10^9 c.f.u. per 7 μ l TSB with or without various agents, placed at the nose tip, and allowed to be inhaled slowly by 7-day-old mice. The exact bacterial concentration in the inoculum was determined by immediately plating out dilutions of the initial inoculum. Infected animals were returned to their cages and observed for 20 h. To quantify infection, the lung and spleen were isolated, weighed, strained by passing through a stainless steel mesh, and incubated for 30 min in TSB containing 0.1% Triton X-100. We diluted the resulting homogenate and plated it onto TSB agar plates to quantify colonized bacteria directly.

Measurement of shed syndecan ectodomains in lung

The relative amounts of soluble shed syndecan-1 and -4 ectodomains were quantified by isolating lungs of 7-day-old mice that had been inoculated with either 7 μ l TSB (control), 2×10^9 c.f.u. per 7 μ l *P. aeruginosa* (PAO1), or 500 ng per 7 μ l of purified LasA in the absence or presence of 20 μ M BB1101. The lungs were strained through a stainless-steel mesh in ice-cold TBS containing 20 mM EDTA, 100 μ M antipain, 30 μ M bestatin, 30 μ M chymostatin, 170 μ M E64, 20 μ M leupeptin, 15 μ M pepstatin, 100 μ M phosphoramidon and 2 μ M aprotinin, and centrifuged at 1,000g and then 100,000g to obtain soluble components. We estimated the protein concentration of the resulting supernatant by the Micro BCA (Pierce) method, and dot-blotted 500 μ g of each sample onto a cationic PVDF membrane (Immobilon N, Millipore). The relative amounts of soluble syndecan-1 and -4 ectodomains were determined by an immuno-detection method using rat anti-mouse syndecan-1 (281-2) and syndecan-4 (Ky 8.2) monoclonal antibodies¹¹. The 281-2 antibody was from Pharmingen (San Diego, CA); the Ky 8.2 monoclonal was a gift from P. Kincade (Oklahoma Medical Research Foundation, Oklahoma City, OK).

Received 20 November 2000; accepted 26 February 2001.

- Park, P. W., Reizes, O. & Bernfield, M. Cell surface heparan sulfate proteoglycans: selective regulators of ligand-receptor encounters. *J. Biol. Chem.* **275**, 29923–29926 (2000).
- Bernfield, M. et al. Functions of cell surface heparan sulfate proteoglycans. *Annu. Rev. Biochem.* **68**, 729–777 (1999).
- Shieh, M. T., WuDunn, D., Montgomery, R. L., Esko, J. D. & Spear, P. G. Cell surface receptors for herpes simplex virus are heparan sulfate proteoglycans. *J. Cell Biol.* **116**, 1273–1281 (1992).
- Shukla, D. et al. A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry. *Cell* **99**, 13–22 (1999).
- Chen, Y. et al. Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat. Med.* **3**, 866–871 (1997).
- Pancake, S. J., Holt, G. D., Mellouk, S. & Hoffman, S. L. Malaria sporozoites and circumsporozoite proteins bind specifically to sulfated glycoconjugates. *J. Cell Biol.* **117**, 1351–1357 (1992).
- Ortega-Barria, E. & Pereira, M. E. A novel *T. cruzi* heparin-binding protein promotes fibroblast adhesion and penetration of engineered bacteria and trypanosomes into mammalian cells. *Cell* **67**, 411–421 (1991).
- Menozi, F. D. et al. Identification of a heparin-binding hemagglutinin present in Mycobacteria. *J. Exp. Med.* **184**, 993–1001 (1996).
- Grasmé, H. et al. Acidic sphingomyelinase mediates entry of *N. gonorrhoeae* into nonphagocytic cells. *Cell* **91**, 605–615 (1997).
- Rostand, K. S. & Esko, J. D. Microbial adherence to and invasion through proteoglycans. *Infect. Immun.* **65**, 1–8 (1997).
- Park, P. W. et al. Syndecan-1 shedding is enhanced by LasA, a secreted virulence factor of *Pseudomonas aeruginosa*. *J. Biol. Chem.* **275**, 3057–3062 (2000).
- Pier, G. B. et al. Role of mutant CFTR in hypersusceptibility of cystic fibrosis patients to lung infections. *Science* **271**, 64–67 (1996).
- Tang, H., Kays, M. & Prince, A. Role of *Pseudomonas aeruginosa* pili in acute pulmonary infection. *Infect. Immun.* **63**, 1278–1285 (1995).

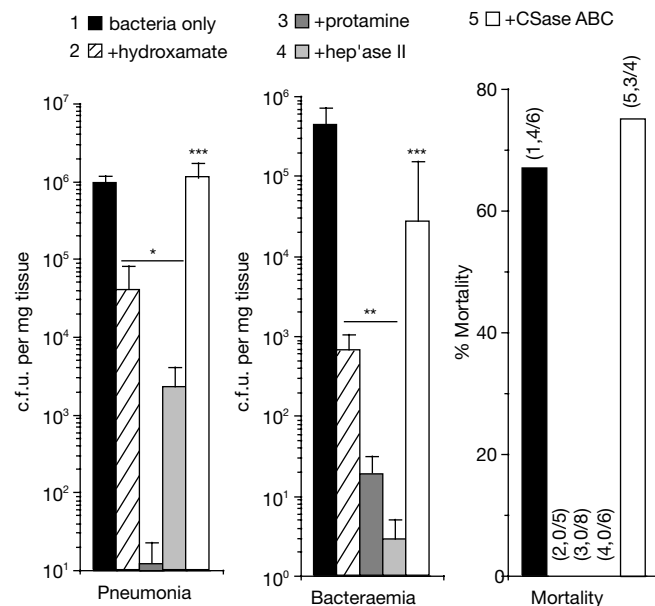


Figure 4 Inhibition of shedding or neutralization of HS prevents intranasally induced *P. aeruginosa* lung infection. Wild-type synd-1^{+/+} mice (7-day-old) were intranasally infected with *P. aeruginosa* strain PAO1 ($(2.0\text{--}2.2) \times 10^9$ c.f.u. per 7 μ l) in the absence or presence of a peptide hydroxamate inhibitor of shedding, BB1101 (20 μ M), or the HS neutralizing agent protamine (200 ng per 7 μ l), and the extent of infection was assessed as in Fig. 1. Alternatively, synd-1^{+/+} mice were pre-treated with 0.3 mU per 7 μ l of heparitinase II (hep'ase II) or chondroitinase ABC (CSase ABC) for 1 h and then inoculated with PAO1 with a fresh dose of 0.3 mU heparitinase II or chondroitinase ABC. Results for pneumonia and bacteraemia represent mean $\pm$ s.e. (* $P \leq 0.02$, ** $P \leq 0.08$, *** $P > 0.34$; Student's *t*-test). Comparison of mortality percentages, synd-1^{-/-} mice given only bacteria versus mice given bacteria and the following reagents: plus hydroxamate (BB1101), $P = 0.045$; plus protamine, $P = 0.015$; plus heparitinase II, $P = 0.03$ (Fisher's exact test).

14. Allewelt, M., Coleman, E. T., Grout, M., Priebe, G. P. & Pier, G. B. Acquisition of expression of the *Pseudomonas aeruginosa* ExoU cytotoxin leads to increased bacterial virulence in a murine model of acute pneumonia and systemic spread. *Infect. Immun.* **68**, 3998–4004 (2000).
15. Alexander, C. M. *et al.* Syndecan-1 is required for Wnt-1-induced mammary tumorigenesis in mice. *Nature Genet.* **25**, 329–332 (2000).
16. Rostand, K. S. & Esko, J. D. Cholesterol and cholesterol esters: host receptors for *Pseudomonas aeruginosa* adherence. *J. Biol. Chem.* **268**, 24053–24059 (1993).
17. Kato, M. *et al.* Physiological degradation converts the soluble syndecan-1 ectodomain from an inhibitor to a potent activator of FGF-2. *Nature Med.* **4**, 691–697 (1998).
18. Subramanian, S. V., Fitzgerald, M. L. & Bernfield, M. Regulated shedding of syndecan-1 and -4 ectodomains by thrombin and growth factor activation. *J. Biol. Chem.* **272**, 14713–14720 (1997).
19. Woods, D. E., Cryz, S. J., Friedman, R. L. & Iglewski, B. H. Contribution of toxin A and elastase to virulence of *Pseudomonas aeruginosa* in chronic lung infections of rats. *Infect. Immun.* **36**, 1223–1228 (1982).
20. Blackwood, L. L., Stone, R. M., Iglewski, B. H. & Pennington, J. E. Evaluation of *Pseudomonas aeruginosa* exotoxin A and elastase as virulence factors in acute lung infection. *Infect. Immun.* **39**, 198–201 (1983).
21. Huttner, K. M. & Bevins, C. L. Antimicrobial peptides as mediators of epithelial host defense. *Pediatr. Res.* **45**, 785–794 (1999).
22. Bals, R., Weiner, D. J. & Wilson, J. M. The innate immune system in cystic fibrosis lung disease. *J. Clin. Invest.* **103**, 303–307 (1999).
23. Kainulainen, V., Wang, H., Schick, C. & Bernfield, M. Syndecans, heparan sulfate proteoglycans, maintain the proteolytic balance of acute wound fluids. *J. Biol. Chem.* **273**, 11563–11569 (1998).
24. Belaouaj, A. *et al.* Mice lacking neutrophil elastase reveal impaired host defense against Gram negative bacterial sepsis. *Nature Med.* **4**, 615–618 (1998).
25. LeVine, A. M. *et al.* Surfactant protein-A-deficient mice are susceptible to *Pseudomonas aeruginosa* infection. *Am. J. Respir. Cell Mol. Biol.* **19**, 700–708 (1998).
26. Crouch, E. C. Collectins and pulmonary host defense. *Am. J. Respir. Cell Mol. Biol.* **19**, 177–201 (1998).
27. Kuschert, G. S. V. *et al.* Glycosaminoglycans interact selectively with chemokines and modulate receptor binding and cellular responses. *Biochemistry* **38**, 12959–12968 (1999).
28. Pearson, J. P., Feldman, M., Iglewski, B. H. & Prince, A. *Pseudomonas aeruginosa* cell-to-cell signaling is required for virulence in a model of acute pulmonary infections. *Infect. Immun.* **68**, 4331–4334 (2000).
29. Schmidtchen, A., Frick, I. & Björck, L. Dermatan sulphate is released by proteinases of common pathogenic bacteria and inactivates antibacterial alpha-defensin. *Mol. Microbiol.* **39**, 708–713 (2001).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as a paper copy from the London editorial office of Nature.

Acknowledgements

We thank A. Drummond and P. Kincade for the peptide hydroxamate (BB1101) and the Ky 8.2 monoclonal antibody, respectively. This work was supported by the Parker B. Francis Foundation (P.W.P.) and the NIH (G.B.P. and M.B.).

Correspondence and requests for materials should be addressed to M.B. (e-mail: merton.bernfield@tch.harvard.edu) or P.W.P. (email: pwpark@bcm.tmc.edu).

Initiation of a G2/M checkpoint after ultraviolet radiation requires p38 kinase

Dmitry V. Bulavin*, Yuichiro Higashimoto†, Ian J. Popoff‡, William A. Gaarde‡, Venkatesha Basrur†, Olga Potapova§, Ettore Appella† & Albert J. Fornace Jr*

* Division of Basic Science and † Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

‡ Isis Pharmaceuticals Inc., Carlsbad, California 92008, USA

§ Laboratory of Biological Chemistry, National Institute on Aging, NIH, Baltimore, Maryland 21224, USA

Response to genotoxic stress can be considered as a multistage process involving initiation of cell-cycle arrest and maintenance of arrest during DNA repair¹. Although maintenance of G2/M checkpoints is known to involve Chk1, Chk2/Rad53 and upstream components, the mechanisms involved in its initiation are less well defined^{1–4}. Here we report that p38 kinase has a critical role in the initiation of a G2 delay after ultraviolet radiation. Inhibition of p38 blocks the rapid initiation of this checkpoint in both

human and murine cells after ultraviolet radiation. *In vitro*, p38 binds and phosphorylates Cdc25B at serines 309 and 361, and Cdc25C at serine 216; phosphorylation of these residues is required for binding to 14-3-3 proteins. *In vivo*, inhibition of p38 prevents both phosphorylation of Cdc25B at serine 309 and 14-3-3 binding after ultraviolet radiation, and mutation of this site is sufficient to inhibit the checkpoint initiation. In contrast, *in vivo* Cdc25C binding to 14-3-3 is not affected by p38 inhibition after ultraviolet radiation. We propose that regulation of Cdc25B phosphorylation by p38 is a critical event for initiating the G2/M checkpoint after ultraviolet radiation.

Evidence exists for important differences in the induction of G2/M delays after DNA strand damage caused by ionizing radiation (IR) and ultraviolet (UV) radiation^{1,5}. Several mechanisms are involved in regulating G2/M delay after IR, and a key component includes activation of Chk1 and Chk2 by the cell cycle proteins ATM and/or ATR, which results in phosphorylation of Cdc25C at Ser 216 (refs 1, 2, 6, 7). Less is known about checkpoint regulation after UV radiation, but evidence indicates that Cdc25 may be involved^{1,7,8}, whereas ATM seems to be involved primarily in IR- rather than UV-induced damage signalling¹. After either IR or UV radiation the initiation of the G2/M checkpoint is rapid, but the regulation of these early events has not been well characterized. Cdc25B, a cytoplasmic protein in G2/M⁹, has been termed a 'starter protein' for G2/M progression as it is activated before Cdc25C^{7,10}; thus, Cdc25B is a candidate for G2 checkpoint initiation.

The mitogen-activated protein kinase (MAPK) pathway has a central role in cellular signalling and two of its three components, the p38 kinase and Jun amino-terminal kinase (JNK), are activated rapidly by many stresses including UV radiation. To address the potential role of p38 kinase in the regulation of cell-cycle progression, we determined the effect of a specific p38 inhibitor (SB202190) on UV-induced checkpoint activation. Whereas inhibition of p38 had no discernible effect on G1 checkpoint activation as measured by progression into S phase (data not shown), the rapid onset of G2 checkpoint was strongly inhibited. For primary isogenic cells from wild-type, p53^{-/-} and p21^{-/-} mice, the onset of G2 checkpoint was inhibited markedly at both low and high UV radiation doses (Fig. 1a); the only exception was weaker but significant ($P < 0.05$) inhibition in p53^{-/-} cells at the lower dose.

G2 checkpoint activation was measured by the reduction in the mitotic index at early times after irradiation⁵; the duration of G2/M in these dermal fibroblasts was estimated to be about 4.5 h, and thus excluded a contribution by UV-induced S-phase delay during the first 4 h. Using 20 J m⁻², inhibition was also seen in p53 wild-type and p53-deficient human tumour lines including HeLa cells (Fig. 2b). Treatment with SB202190 reduced p38 activity in all three types of dermal fibroblast, as determined by phosphorylation of glutathione S-transferase (GST)–ATF2, whereas it had no effect on JNK activity, as measured by GST–Jun phosphorylation (Fig. 1a). Neither the inhibitor nor UV radiation affected the level of cellular p38 (data not shown). Notably, changes in the mitotic index correlated with Cdc2 kinase activity, as measured by phosphorylation of histone H1 (Fig. 1a, HH1). Ultraviolet radiation decreased Cdc2 kinase activity in all cell lines tested, and p38 inhibition completely blocked the decrease during the first hours after irradiation.

The role for p38 in the control of G2 progression seems to be specific for particular types of stress, as it affected neither IR-induced checkpoint activation nor normal G2/M progression. Ionizing radiation frequently does not induce appreciable activation of MAPKs including p38 (ref. 11), and the p38 inhibitor had no effect on the rapid onset of G2 checkpoint in either wild-type or p53-deficient cells (Fig. 1b). Moreover, inhibitors of IR signalling proteins, such as caffeine (ATM/ATR) and UCN-01 (Chk1), had no effect on the initiation of the UV-induced G2 checkpoint (Fig. 2b). In non-irradiated cells, inhibition of p38 also had no detectable

effect on normal G2/M progression (Fig. 1c).

The p38-mediated G2 checkpoint occurs in p53-deficient cells, but seemed to be attenuated at the lower UV dose (Fig. 1a). As higher doses triggered strong activation of the p38-dependent checkpoint regardless of p53 status, we carried out further studies in p53-deficient HeLa cells. HeLa cells express only two of the four isoforms of p38 at appreciable levels¹²; both p38 α and p38 β are strongly inhibited by SB202190, whereas the other two p38 isoforms are not¹³. We could selectively block these two p38 isoforms, which are ubiquitously expressed in most cell types, with an antisense oligonucleotide approach (Fig. 2a) that also attenuated UV-induced checkpoint activation. Suppression of either p38 α or p38 β levels interfered with normal UV checkpoint activation with slightly different kinetics (Fig. 2b). The p38 α oligonucleotide could reduce G2 arrest to that achieved with SB202190, whereas inhibition of p38 β had a more transient effect with prominent attenuation only in the first hour after irradiation. These results support a role for both p38 isoforms in checkpoint activation.

We attempted to identify the p38 substrate(s) involved in G2 checkpoint control. Previous studies have indicated that a peptide containing an LSP motif is a substrate for p38, and this motif can be found in many proteins including the cell-cycle control proteins p53 (Ser 33 and Ser 46)¹⁴, cyclin B1 (Ser 116), Myt1 (Ser 431) and

Cdc25B (Ser 15, Ser 224, Ser 235). Mutation of Ser 116 had no effect on cyclin B1 localization or on Cdc2/cyclin B1 kinase activity (data not shown). Likewise, mutation of Ser 431 to alanine in Myt1 did not affect its ability to inhibit Cdc2 kinase activity, nor did it affect the ability of Myt1 to sequester either wild-type or S116A cyclin B1 in the cytoplasm when nuclear export of cyclin B1 was interrupted by leptomycin B (data not shown). Ultraviolet radiation did not affect the cellular localization or protein level of Cdc25B (data not shown). In contrast, it has been shown that incubating Cdc2 complex from UV-irradiated cells with GST-Cdc25B is sufficient to restore Cdc2 activity¹⁵, indicating that Cdc2 phosphorylation is a more likely target for G2 control after UV radiation.

Cdc25B activity is essential for G2/M transition¹⁰ and has been implicated in the regulation of Cdc2 activity after UV radiation⁸. Consequently, we synthesized peptides corresponding to regions containing the LSP motif, as well as to the 14-3-3-binding region (303–312), and incubated them with p38 α -Flag isolated from UV-irradiated cells¹⁴. Unexpectedly, p38 α phosphorylated the 14-3-3-binding region at least 25 times more efficiently than the LSP peptides (Fig. 3a, lanes 1–4). We next used peptides that matched the consensus 14-3-3-binding sites¹⁶ (RSXpSXP or RXY/FXpSXP) in Cdc25B (containing Ser 137, Ser 216, Ser 309 and Ser 361) and

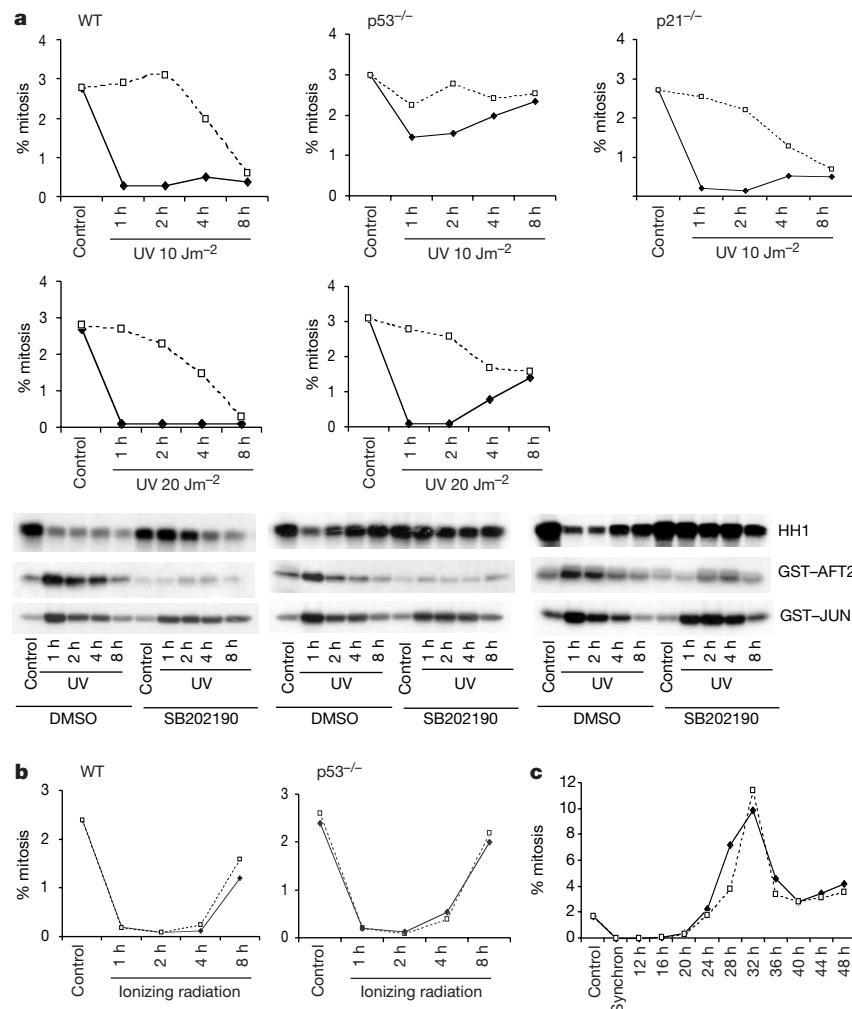
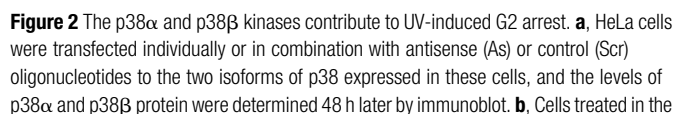


Figure 1 A principal role for p38 kinase in G2 checkpoint activation after UV radiation.

a, Dermal fibroblasts (DFs) established from wild-type, p53^{-/-} and p21^{-/-} mice were UV irradiated in the presence of 10 μ M p38 inhibitor SB202190 (open symbols) or dimethyl sulphoxide (DMSO; filled symbols) as a control, and the percentage of cells in mitosis and Cdc2, p38 and JNK activities were determined. **b**, Wild-type and p53^{-/-} DFs were

γ -irradiated (3 Gy) in the presence of SB202190 or DMSO and the mitotic indexes were determined as in **a**. **c**, Wild-type DFs were synchronized with 0.25% serum for 3 d, and then stimulated with 20% serum. SB202190 or DMSO was added 20 h later and the mitotic index was determined.



a

→
→
→

10-20
219-288
231-241
303-312
303-312 (S307A)
303-312 (S309A)
132-141
210-219
356-365
356-365 (S359A)
356-365 (S361A)
210-220
210-220 (S214A)
210-220 (S216A)

Cdc25B Cdc25C

b

Phosphorylation rate 1 1 0.8 0.6 0.8 0.6 0.1 0.1 1 0.1

GST Cdc25B WT Cdc25B S137A Cdc25B S309A Cdc25B S361A Cdc25B S137/309A Cdc25B S137/361A Cdc25B S309/361A Cdc25B S137/309/361A Cdc25C WT Cdc25C S216A

32P Cdc25B Cdc25C

c

p38

Cdc25B Cdc25C GST

GST GST-Cdc25B GST-Cdc25C

d

14-3-3

Cdc25B Cdc25C

GST GST+p38 Cdc25B WT Cdc25B WT+p38 Cdc25B S309A Cdc25B S309A+p38 Cdc25B S361A Cdc25B S361A+p38 Cdc25B S309/361A Cdc25B S309/361A+p38 Cdc25C WT Cdc25C WT+p38 Cdc25C S216A Cdc25C S216A+p38

either wild-type or the designated mutant of Cdc25B and Cdc25C. **c**, Association of p38 α kinase with Cdc25B and Cdc25C. **d**, Association of p38-phosphorylated Cdc25 with 14-3-3 was analysed as described in Methods.

Cdc25C (containing Ser 216). All these sites, with the exception of Cdc25B Ser 216, were phosphorylated (Fig. 3a).

We extended these experiments to whole proteins and found that mutation of Ser 216 in Cdc25C or the combination of Ser 309/Ser 361 in Cdc25B reduced phosphorylation by p38 to about 10% of that in the wild-type protein (Fig. 3b). Mutation at Ser 137 either alone or in combination with Ser 309 or Ser 361 had no appreciable effect, and thus we conclude that Ser 309/Ser 361 in Cdc25B and Ser 216 in Cdc25C are the most critical sites for phosphorylation by p38 α and p38 β (data not shown) kinases.

To determine whether p38 kinase and the Cdc25 proteins form a complex, GST–Cdc25C and GST–Cdc25B bound to glutathione beads were incubated with cell extracts from HeLa cells and precipitated. Both fusion proteins bound p38 kinase specifically, whereas GST alone did not (Fig. 3c), indicating the presence of a complex. Notably, mutation of the phosphorylation sites on both proteins, as well as UV irradiation of HeLa cells, did not affect binding with p38 kinase (data not shown).

Phosphorylation of Cdc25 triggers cell-cycle arrest by the sequestration of Cdc25 by 14-3-3 (refs 1, 17, 18). We found that phosphorylation of Cdc25B or Cdc25C by p38 kinase results in a significant increase of 14-3-3 binding (Fig. 3d). For Cdc25C, this increased binding was primarily dependent on phosphorylation of Ser 216, because mutation of this site significantly reduced p38-kinase-induced binding to 14-3-3 proteins. For Cdc25B, mutation of either Ser 309 or Ser 361 alone was sufficient to reduce 14-3-3 binding to that observed with the S309A/S361A double mutant (Fig. 3d); this indicates that phosphorylation at both sites is required for strong p38-kinase-induced binding.

Both Cdc25B and Cdc25C bound to 14-3-3 under normal

conditions *in vivo* and inhibition of p38 did not affect this binding, whereas mutation of Ser 309 in Cdc25B or Ser 216 in Cdc25C to alanine decreased complex formation (Fig. 4a). Ultraviolet radiation had no significant effect on 14-3-3 binding, nor did inactivation of p38 change complex formation with Cdc25C after UV radiation. For Cdc25B, however, 14-3-3 binding was significantly reduced in UV-irradiated cells after p38 inactivation (Fig. 4a). These results support an important role for p38 in the regulation of Cdc25B by 14-3-3 after stresses such as UV radiation, but argue against a significant role in the regulation of Cdc25C.

To address further the *in vivo* role for p38 kinase in regulating phosphorylation of Cdc25B Ser 309, we generated a phospho-specific antibody to this site. Analysis of non-synchronized HeLa cells showed that this site is phosphorylated in non-synchronized cells as well as in G1, S and G2 phases of the cell cycle (Fig. 4b). In mitosis, however, Cdc25B phosphorylation at Ser 309 was reduced markedly, correlating with high Cdc25B phosphatase activity (Fig. 4b, bottom panel). Ultraviolet radiation had no effect on Cdc25B phosphorylation at Ser 309, and inactivation of p38 kinase did not affect Ser 309 phosphorylation in control cells, but this phosphorylation was significantly reduced in asynchronous UV-irradiated cells in the presence of the p38 inhibitor (Fig. 4c).

We investigated the role of Cdc25B Ser 309 in regulating the initiation of G2 checkpoint. Synchronized HeLa cells were transfected with different green fluorescent protein (GFP) fusion constructs, and progression through G2 phase was analysed (Fig. 4d). Under these conditions, cells entered mitosis 8–9 h after release in complete medium and only 3% of the GFP-positive cells could reach mitosis by 7 h (Fig. 4d). As overexpression of wild-type Cdc25B induces premature onset of mitosis (Fig. 4b)⁹, the cells were UV

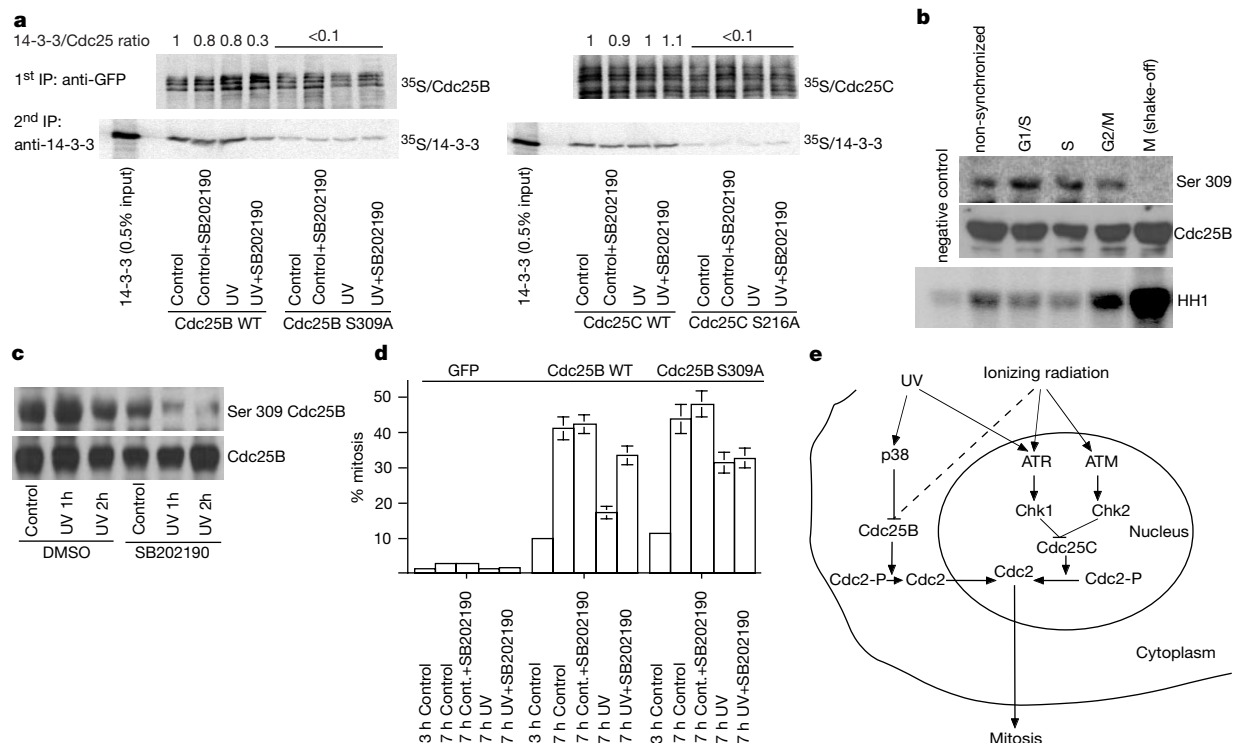


Figure 4 *In vivo* phosphorylation of Ser 309 Cdc25B after UV radiation, and its role in regulating 14-3-3 binding and G2/M checkpoint control. **a**, The level of 14-3-3 in complex with Cdc25 *in vivo* was determined after sequential immunoprecipitations (see Methods). **b**, Ser 309 Cdc25B phosphorylation and Cdc25B activity (bottom panel) were analysed at the different phases of the cell cycle. **c**, HeLa cells were treated with DMSO or SB202190 (20 μ M) and UV irradiated (20 J m⁻²) after 30 min; Ser 309 Cdc25B phosphorylation was then analysed. **d**, In cell synchrony experiments, cells were

transfected with the indicated GFP fusion plasmids (see Methods). Cells, which had progressed past G2 phase, were arrested in mitosis with nocodazole over the next 4 h (7 h control, unirradiated samples; 7 h control + SB202190, unirradiated samples in the presence of p38 inhibitor; 7 h UV, UV-irradiated cells; 7 h UV + SB202190, UV-irradiated cells in the presence of the p38 inhibitor). The number of GFP-positive cells in mitosis was counted. **e**, A model for the role of p38 kinase in regulating G2/M checkpoint activation after genotoxic stress. Dotted line refers to unidentified, activated kinase.

irradiated only 3 h after release in complete medium. Ultraviolet radiation abrogated G2 progression in Cdc25B wild-type GFP-positive cells, and the p38 inhibitor could overcome this checkpoint. The Cdc25B S309A mutant also induced premature entry into mitosis, but UV radiation had much less effect on G2 progression, and inclusion of the p38 inhibitor had no significant effect on G2 progression after UV radiation (Fig. 4d). These data show that abrogation of Cdc25B Ser 309 phosphorylation is sufficient to overcome initiation of the G2 checkpoint after UV radiation, and that this site is a critical target for p38 kinase in its regulation of checkpoint initiation.

Before mitosis, Cdc25B and Cdc25C are phosphorylated at Ser 309 and Ser 216, respectively, with resultant complex formation with 14-3-3 proteins. Maintaining this phosphorylation *in vivo* must involve one or more kinases, which regulate entry into mitosis under normal (non-stress) conditions, although the specific kinases and how their activities are controlled remain uncertain. After genotoxic stress Cdc25 phosphorylation is required to prevent progression into mitosis, but the exact mechanism has not been certain because these inhibitory sites are already phosphorylated in G2 phase, even in unstressed cells.

On the basis of our results and data for Chk1-dependent phosphorylation of Cdc25C after IR¹⁹, a model can be proposed for the activation of the G2 checkpoint (Fig. 4e). After genotoxic stress, cells activate a mechanism that switches the system of kinases regulating Cdc25 phosphorylation in unstressed cells to kinases that are stress-inducible. By switching kinases after stress, the cell creates a system that becomes 'a sensor' for damage and allows entry to mitosis only after appropriate stress recovery. In this case, p38 kinase is an early 'sensor' of cell damage because it regulates Cdc25B phosphorylation (Fig. 4c). Later events in the maintenance of G2 arrest involve other components (such as ATR/ATM pathways) that participate in regulating Cdc25C in the nucleus and can be induced by most stresses including UV radiation as well as IR. For example, disruption of either the Chk1 or Chk2 genes attenuated only the later time points (12 h and later) but not the initiation of the UV or IR G2/M checkpoints^{4,20}. It is striking that both initiation and maintenance pathways target the same system of Cdc25 phosphatases, emphasizing their central role in the regulation of G2/M progression after genotoxic stress.

Note added in proof: It has recently been shown that a yeast analogue of p38 kinase, HOG1, is required for hypertonic stress-induced G2 arrest²³. □

Methods

Plasmids and site-directed mutagenesis

p38 α and p38 β expression vectors were obtained from R. Davis and J. Han, respectively. GST-Cdc25B and GST-Cdc25C plasmids were provided by I. Hoffmann and H. Piwnicka-Worms. We obtained GFP-Cdc25B and GFP-Cdc25C from J. Pines. To generate site mutants, we used a site-directed mutagenesis kit (Stratagene) with the following pairs of primers: for Cdc25B S137A, 5'-cagacgtctcaggtatgccggtgaggtgc-3'/5'-gcagcctaccggcatagcctggaagcgtctg-3'; S309A, 5'-cggctcttcgctctccggcatgcctgcag-3'/5'-ctgcagggcatggccggagagcggaagacgc-3'; S361A, 5'-gtcctccgctcaaaagcactgtgcatgag-3'/5'-ctcatctgacacagtgctttgagcggaggac-3'; and for Cdc25C S216A, 5'-cctatatcgctcccgccg atgcagagaaac-3' / 5'-gttctctggcatgcggggagcgatag-3'. Results were verified by DNA sequencing.

Analysis of protein binding

Bacterially expressed GST proteins (GST, GST-Cdc25B, GST-Cdc25C or the indicated mutants) on glutathione beads were incubated overnight with protein extracts from HeLa cells (1 mg) and precipitated. p38 α was analysed by C-20, and p38 β by C-16 polyclonal antibodies (Santa Cruz). To analyse 14-3-3 binding, GST proteins were phosphorylated for 1 h with UV-activated p38 α (ref. 15); we used a UVC radiation source for all experiments. 14-3-3 proteins in complex with GST proteins were analysed by immunoblotting with polyclonal antibody (K-19, Santa Cruz). For analysis of Cdc25 interaction with 14-3-3 proteins, HeLa cells were co-transfected with 5 μ g of GFP-tagged Cdc25 plasmids and 5 μ g of haemagglutinin A (HA)-tagged 14-3-3. On the next day, cells were labelled with 300 μ Ci [³⁵S]methionine per ml of EasyTag Express Protein Labeling Mix (NEN) for 60 min and then UV irradiated (60 J m⁻²). For some samples, we added 40 μ M SB202190 30 min before UV radiation to block p38 activity. Labelling was continued for

the next 60 min, and the cells were gathered. Cdc25s were precipitated with anti-GFP antibody. After boiling in SDS sample buffer, precipitates were diluted 30 times with lysis buffer, and 14-3-3 was subsequently re-immunoprecipitated with 10 μ l of anti 14-3-3 antibody (Santa Cruz) and 5 μ l of anti-HA antibody (Babco).

Kinase and phosphatase assays

The p38 and Cdc2 kinase assays were performed after immunoprecipitation with specific antibodies^{14,21} using GST-ATF2 and histone H1, respectively. We analysed JNK kinase after pull down with GST-Jun. Cdc25B activity was determined after dephosphorylation of Cdc2, with subsequent analysis of Cdc2 activity using histone H1 as a substrate²¹.

Analysis of the role of Ser 309 Cdc25B

HeLa cells were synchronized by a double-thymidine block procedure as described²². Two hours before the second incubation with thymidine, we transfected the cells with the indicated GFP fusion plasmids (Fig. 4d). Cells were irradiated 3 h after release into complete medium with 20 J m⁻², and nocodazole was added during the next 4 h to trap any cells that had progressed through G2 into mitosis. We determined the number of GFP-positive cells in mitosis as described⁵. Mitotic cells were obtained by a mitotic shake-off procedure: cells were collected at 8 h after release from the double-thymidine block, when the fraction of G2/M cells was maximal as determined by FACS, and then every 45 min thereafter during the next 2.25 h.

Analysis of phosphorylation

Ser 216 Cdc25C antibody was from Cell Signaling Technology. Rabbit polyclonal antibody specific for phosphorylation at Ser 309 Cdc25B was raised against the sequence Ac-305-314(309P) (that is, Ac-FRSPS(P)-MPCSV) conjugated to keyhole limpet haemocyanin. We confirmed the specificity of antibody by enzyme-linked immunoabsorbent and immunoblot assays with GST proteins.

Oligonucleotide synthesis and treatment

The 20-nucleotide oligonucleotides used in this study were synthesized at ISIS Pharmaceuticals Inc. as follows: p38 α antisense, 5'-ttctctatctagtgccaat-3'; p38 α scrambled, 5'-ttatcctagcttagacat-3'; p38 β antisense, 5'-gtacgttctgcgcgcgtgga-3'; p38 β scrambled, 5'-gttcgctggcgtctgtgca-3'. All oligonucleotides contained five 2'-O-methoxyethyl-modified sugar/phosphodiester residues at the 5' and 3' ends, flanking the centre ten 2'-deoxyoligonucleotide/phosphorothioate residues. Cells were treated with 0.3 μ M oligonucleotides in the presence of 10 μ g ml⁻¹ of Lipofectin reagent (Life Technologies).

Received 4 December 2000; accepted 23 February 2001.

- Pines, J. Four-dimensional control of the cell cycle. *Nature Cell Biol.* **1**, E73–E79 (1999).
- Peng, C. Y. *et al.* Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216. *Science* **277**, 1501–1505 (1997).
- Furnari, B., Rhind, N. & Russell, P. Cdc25 mitotic inducer targeted by chk1 DNA damage checkpoint kinase. *Science* **277**, 1495–1497 (1997).
- Hirao, A. *et al.* DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science* **287**, 1824–1827 (2000).
- Hollander, M. C. *et al.* Genomic instability in *Gadd45a*-deficient mice. *Nature Genet.* **23**, 176–184 (1999).
- Sanchez, Y. *et al.* Conservation of the Chk1 checkpoint pathway in mammals: linkage of DNA damage to Cdk regulation through Cdc25. *Science* **277**, 1497–1501 (1997).
- Nilsson, I. & Hoffmann, I. Cell cycle regulation by the Cdc25 phosphatase family. *Prog. Cell Cycle Res.* **4**, 107–114 (2000).
- Gabrieli, B. G., Clark, J. M., McCormack, A. K. & Ellem, K. A. Ultraviolet light-induced G2 phase cell cycle checkpoint blocks cdc25- dependent progression into mitosis. *Oncogene* **15**, 749–758 (1997).
- Karlsson, C., Katich, S., Hagting, A., Hoffmann, I. & Pines, J. Cdc25B and Cdc25C differ markedly in their properties as initiators of mitosis. *J. Cell. Biol.* **146**, 573–584 (1999).
- Lammer, C. *et al.* The cdc25B phosphatase is essential for the G2/M phase transition in human cells. *J. Cell Sci.* **111**, 2445–2453 (1998).
- Liu, Z. G. *et al.* Three distinct signalling responses by murine fibroblasts to genotoxic stress. *Nature* **384**, 273–276 (1996).
- Jiang, Y. *et al.* Characterization of the structure and function of the fourth member of p38 group mitogen-activated protein kinases, p38 δ . *J. Biol. Chem.* **272**, 30122–30128 (1997).
- Lee, J. C., Kassis, S., Kumar, S., Badger, A. & Adams, J. L. p38 mitogen-activated protein kinase inhibitors—mechanisms and therapeutic potentials. *Pharmacol. Ther.* **82**, 389–397 (1999).
- Bulavin, D. *et al.* Phosphorylation of human p53 by p38 kinase coordinates N-terminal phosphorylation and apoptosis in response to UV radiation. *EMBO J.* **18**, 6845–6854 (1999).
- Poon, R. Y. C., Jiang, W., Toyoshima, H. & Hunter, T. Cyclin-dependent kinases are inactivated by a combination of p21 and Thr-14/Tyr-15 phosphorylation after UV-induced DNA damage. *J. Biol. Chem.* **271**, 13283–13291 (1996).
- Yaffe, M. B. *et al.* The structural basis for 14-3-3:phosphopeptide binding specificity. *Cell* **91**, 961–971 (1997).
- Lopez-Girona, A., Furnari, B., Mondesert, O. & Russell, P. Nuclear localization of Cdc25 is regulated by DNA damage and a 14-3-3 protein. *Nature* **397**, 172–175 (1999).
- Morris, M. C., Heitz, A., Mery, J., Heitz, F. & Divita, G. B. An essential phosphorylation-site domain of human CDC25C interacts with both 14-3-3 and cyclins. *J. Biol. Chem.* **275**, 28849–28857 (2000).
- Graves, P. R. *et al.* The Chk1 protein kinase and the Cdc25C regulatory pathways are targets of the anticancer agent UCN-01. *J. Biol. Chem.* **275**, 5600–5605 (2000).
- Liu, Q. *et al.* Chk1 is an essential kinase that is regulated by Atr and required for the G2/M DNA damage checkpoint. *Genes Dev.* **14**, 1448–1459 (2000).

21. Blasina, A. *et al.* A human homologue of the checkpoint kinase Cds1 directly inhibits Cdc25 phosphatase. *Curr. Biol.* **9**, 1–10 (1999).
22. Liu, F., Rothblum-Oviatt, C., Ryan, C. E. & Piwnicka-Worms, H. Overproduction of human Myt1 kinase induces a G2 cell cycle delay by interfering with the intracellular trafficking of Cdc2-cyclin B1 complexes. *Mol. Cell. Biol.* **19**, 5113–5123 (1999).
23. Alexander, M. R. *et al.* Regulation of cell cycle progression by *swe1p* and *hog1p* following hypertonic stress. *Mol. Biol. Cell* **12**, 53–62 (2001).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We are grateful to J. Pines, I. Hoffmann, H. Piwnicka-Worms, J. Han, B. Vogelstein and R. Davis for the plasmids used in the paper; D. Ferris for Cdc2 polyclonal antibody; B. Monia for advice on the antisense oligonucleotide approach; J. Clark for help with DNA sequencing; and O. Kovalsky for help in protein isolation.

Correspondence and requests for materials should be addressed to A.J.F. (e-mail: fornace@nih.gov).

Microarrays of cells expressing defined cDNAs

Junaid Ziauddin & David M. Sabatini

Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

Genome and expressed sequence tag projects are rapidly cataloguing and cloning the genes of higher organisms, including humans. An emerging challenge is to rapidly uncover the functions of genes and to identify gene products with desired properties. We have developed a microarray-driven gene expression system for the functional analysis of many gene products in parallel. Mammalian cells are cultured on a glass slide printed in defined locations with different DNAs. Cells growing on the printed areas take up the DNA, creating spots of localized transfection within a lawn of non-transfected cells. By printing sets of complementary DNAs cloned in expression vectors, we make microarrays whose features are clusters of live cells that express a defined cDNA at each location. Here we demonstrate two uses for our approach: as an alternative to protein microarrays for the identification of drug targets, and as an expression cloning system for the discovery of gene products that alter cellular physiology. By screening transfected cell microarrays expressing 192 different cDNAs, we identified proteins involved in tyrosine kinase signalling, apoptosis and cell adhesion, and with distinct subcellular distributions.

The growing collection of gene sequences and cloned cDNAs demands the development of systematic and high-throughput approaches to characterizing the gene products. Strategies comparable to DNA microarrays for transcriptional profiling^{1,2} and to yeast two-hybrid arrays for determining protein–protein interactions³ do not exist to analyse the function, within mammalian cells, of large sets of genes. At present, *in vivo* gene analysis can be done—on a gene-by-gene scale—by expressing a DNA construct within cells that directs the overproduction of a gene product or inhibits its synthesis or function. The effects on cellular physiology of altering the level of a gene product is then detected using a variety of functional assays. We describe a strategy for the high-throughput analysis of gene function in mammalian cells. We have developed a system suitable for rapidly screening large sets of cDNAs or DNA constructs for those genes encoding desired products or causing cellular phenotypes of interest. Using slides printed with sets of cDNAs in expression vectors, we create living microarrays of cell clusters expressing the gene products. The cell clusters can be

screened for any property detectable on a surface and the identity of the responsible cDNA determined from the coordinates of the cell cluster with a phenotype of interest.

To create these microarrays, we simultaneously transfect distinct and defined areas of a lawn of cells with different plasmid DNAs (Fig. 1a). This is done without the use of individual wells to sequester the DNAs. Nanolitre volumes of plasmid DNA in an aqueous gelatin solution are printed on a glass slide using a robotic arrayer. After drying, the DNA spots are briefly exposed to a lipid transfection reagent; the slide is then placed in a culture dish and covered with adherent mammalian cells in medium. The cells growing on the DNA and gelatin spots express the DNA and divide 2–3 times in the process of creating a microarray with features consisting of clusters of transfected cells (which we call ‘transfected cell microarrays’). We call the method to make the arrays ‘reverse transfection’ because, compared with conventional transfection, we have reversed the order of addition of DNA and adherent cells. For detailed protocols of this and alternative methods see http://staffa.wi.mit.edu/sabatini_public/reverse_transfection.htm.

To illustrate the method, we printed an array with elements containing an expression construct for the green fluorescent protein (GFP). HEK293 cells were plated on the slide for transfection and the fluorescence of the cells detected with a laser fluorescence scanner. A low magnification scan shows a regular pattern of fluorescent spots that matches the pattern in which we printed the GFP expression construct (Fig. 1b). A higher magnification image obtained through fluorescence microscopy shows that each spot is about 120–150 μm in diameter and consists of a cluster of 30–80 fluorescent cells (Fig. 1c). As in a conventional transfection, the total expression level in the clusters is proportional over a

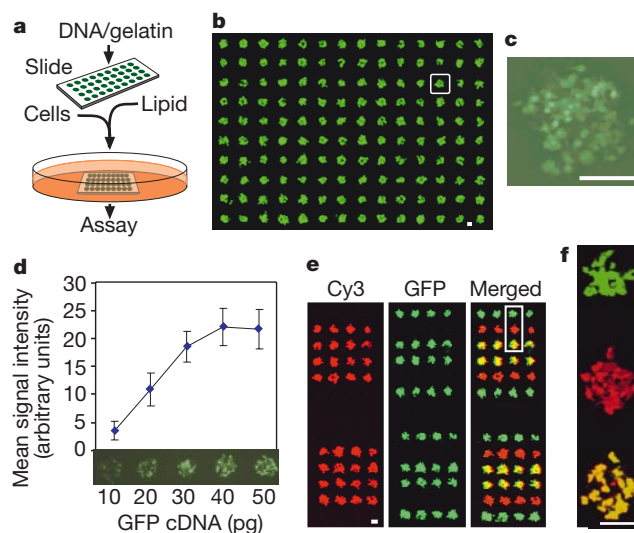


Figure 1 Well-less transfection of plasmid DNAs in defined areas of a lawn of mammalian cells. **a**, Protocol for making microarrays of transfected cells. **b**, Laser scan image of a GFP-expressing microarray made from a slide printed in a 14×10 pattern with a GFP expression construct. **c**, Higher magnification image obtained with fluorescence microscopy of the cell cluster boxed in **b**. Scale bar, 100 μm . **d**, Expression levels of cell clusters in a microarray are proportional, over a fourfold range, to the amount of plasmid DNA printed on the slide. Indicated amounts of the GFP construct assume a 1-nl printing volume. The graph shows the mean $\pm$ s.d. of the fluorescence intensities of the cell clusters ($n = 6$). The fluorescent image is from a representative experiment. **e**, Co-transfection is possible with transfected cell microarrays. Arrays with elements containing expression constructs for HA–GST, GFP or both were transfected and processed for immunofluorescence and imaged with a laser scanner. Cy3, cell clusters expressing HA–GST; GFP, cell clusters expressing GFP; merged, superimposition of Cy3 and GFP signals. Yellow colour indicates co-expression. Scale bar, 100 μm . **f**, Enlarged view of boxed area of scan image from **e**.

defined range to the amount of plasmid DNA used (Fig. 1d). As it is often useful to express two different plasmids in the same cells, we examined whether our technique is compatible with co-transfection. Arrays with elements containing expression constructs for GFP, haemagglutinin (HA)-tagged glutathione S-transferase (GST) or both were prepared and transfected. The cells growing on elements printed with both cDNAs express both encoded proteins, indicating that co-transfection does occur (Fig. 1e).

We first determined whether transfected cell microarrays could be used to identify gene products on the basis of their intrinsic properties. As a test case, we sought to use an array to identify the receptor for FK506, a clinically important immunosuppressant whose pharmacologically relevant target, FKBP12, is an intracellular protein^{4–6}. We printed on a slide, in an easily recognizable pattern, elements containing expression constructs for Myc-tagged FKBP12, GFP or both. After the microarray of transfected cell clusters formed, we added radiolabelled FK506 to the tissue culture medium for 1 h before processing the slide for immunofluorescence and autoradiography. The radiolabelled FK506 bound to the array in a pattern of spots that exactly matched the pattern of cell clusters expressing FKBP12 (Fig. 2a). The GFP-expressing clusters and the non-transfected cells surrounding the clusters had background levels of signal (Fig. 2a). Detection of the bound FK506 with autoradiographical emulsion confirms, at the cellular level, colocalization between FKBP12 expression and FK506 binding (Fig. 2b). The clusters overexpressing FKBP12 had only a 3.1 ± 0.2 -fold increase in the binding of radiolabelled FK506 compared with the GFP-expressing clusters (Fig. 2c), probably reflecting the high levels of endogenous FK506-binding proteins expressed in the cells⁷. Consistent with this, the prior addition of excess rapamycin, a competitive antagonist of FK506, not only eliminated the specific signal but also greatly lowered the background signal (Fig. 2c).

We also detected binding of a small molecule to a membrane

protein expressed in specific clusters of a transfected cell microarray. We transfected slides printed with plasmids encoding an IRES–GFP cassette downstream of the cDNA for the dopamine D1 or serotonin 1A receptor. Cell clusters transfected with either plasmid expressed GFP, but a dopamine antagonist bound only to the cells transfected with the plasmid for the dopamine receptor (Fig. 2d).

We next used transfected cell microarrays to screen a collection, enriched for signalling molecules, of 192 epitope-tagged cDNAs cloned in expression vectors. All screens were performed in a blinded fashion without knowledge of the identity of the encoded proteins. We screened for gene products that increase the activity of kinase signalling pathways, induce apoptosis or are involved in cell–cell adhesion. After transfection of HEK293T cells on the array, we used phosphospecific antibodies to identify cell clusters with increased levels of phosphotyrosine or of the activated phosphorylated forms of several members of the mitogen-activated protein kinase (MAPK) family. Six cell clusters (1A2, 1C7, 1C9, 1C11, 1E3 and 1F6) had phosphotyrosine levels above background (Fig. 3a). The coordinates of the clusters match those of the wells of a microtitre plate containing the source cDNAs, and were used to ascertain the identity of the transfected cDNAs. This revealed that five of these clusters were expressing known tyrosine kinases (TrkC, Syk, Syn, Lck and Blk) whereas the sixth (1C9) encodes a protein of unknown function (Table 1). Screening of duplicate arrays for phospho-MAPKs showed that the clusters expressing Syk and Blk also had increased levels of activated phospho-JNK and phospho-p38 (Table 1). Syk is a known activator of these MAPKs⁸, but this function has not been ascribed previously to Blk. Overexpression of Syn and Lck did not activate JNK or p38, indicating that, at least in HEK293 cells, not all cytoplasmic tyrosine kinases have this effect. To identify gene products that might have a function in apoptosis or cell adhesion, we examined the microarray for clusters containing cells with abnormal morphologies. We found one cluster (2E8) in which the cells were fragmented (Fig. 3b) and positive for the TdT-mediated dUTP nick end-labelling⁹ reaction (data not shown),

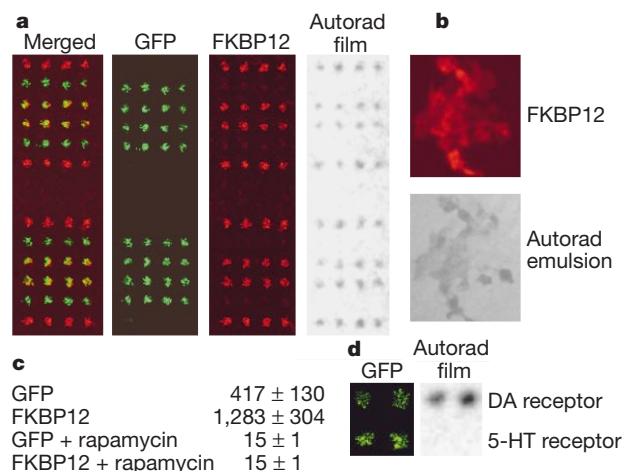


Figure 2 Detection of drug receptors on transfected cell microarray. **a**, Increased binding of tritiated FK506 to cell clusters expressing FKBP12. Laser scans show clusters expressing GFP, Myc-FKBP12 or both (merged). Film autoradiography detects binding of tritiated FK506 to the same array (autorad film). **b**, Colocalization between FKBP12-expressing cells and FK506-binding. Higher magnification image obtained by fluorescence microscopy of an FKBP12-expressing cluster (FKBP12). Emulsion autoradiography detects, with cellular resolution, binding of tritiated FK506 to the same cluster (autorad emulsion). **c**, Rapamycin competes with FK506 for binding to FKBP12-expressing clusters. Mean ± s.d. ($n = 10$) is shown in arbitrary units. **d**, Binding of tritiated SCH23390 to cell clusters expressing the dopamine D1 (DA) but not the serotonin 1A (5-HT) receptor (autorad film). Each receptor was expressed from a plasmid encoding a bicistronic messenger RNA containing an IRES–GFP element downstream of the receptor cDNA. Clusters on the left and right were transfected with 20 and 30 µg plasmid DNA, respectively. Fluorescence microscopy shows that all clusters express GFP.

Table 1 Phenotypes induced and subcellular distributions for cDNAs expressed in transfected cell microarray

Grid: coordinate	Phenotype/distribution	Accession number	Gene name
1:A2	P-Tyr	U05012	TrkC
1:C7	P-Tyr, p-JNK, p-p38	L28824	Syk
1:C9	P-Tyr		Homologous to BT0590
1:C11	P-Tyr	M14333	Syn
1:E3	P-Tyr	M36881	Lck
1:F6	P-Tyr, p-JNK, p-p38	S76617	Blk
1:C2	Cytoskel. changes	M21616	β-PDGF receptor
2:E8	apoptosis	AF016266	TRAIL receptor 2
2:D10	Cell adhesion	X97229	NK receptor
2:F7	Cell adhesion	M98399	CD36
1:A9	Nuclear	U11791	Cyclin H
1:B5	Nuclear	M60527	Deoxycytidine kinase
1:B12	Nuclear	M60724	p70 S6 kinase kinase α
1:C12	Nuclear	M90813	D-type cyclin
1:E4	Mitochondrial	U54645	Methylmalonyl-coA mutase
1:E10	Mitochondrial	J05401	Creatine kinase
1:G9	Nuc/cyto	U40989	Tat interactive protein (TIP60)
1:G10	Nuc/cyto	U09578	3pk kinase
2:A9	Nuclear	X83928	TFIID subunit TAFII28
2:A12	Nuc/cyto	M62831	ETR101
2:B6	Nuc/cyto	X06948	IgE receptor α-subunit
2:B12	Nuclear	X63469	TFIIIE β-subunit
2:C5	Nuclear	M76766	General transcription factor IIB
2:C7	Nuc/cyto	M15059	CD23A
2:C12	Nuclear	X80910	PP1, β-catalytic subunit
2:D4	Nuclear	AF017307	Ets-related transcription factor
2:E7	Nuclear	X63468	TFIIIE-α
2:E12	Nuclear	U22662	Orphan receptor LXR-α
2:F8	Nuclear	L08895	MEF2C
2:F12	Nuclear	AF028008	SP1-like transcription factor
2:G2	Nuc/cyto	U37352	PP2A, regulatory B'α1 subunit

Nuc/cyto, nuclear and cytoplasmic staining was visible. P-Tyr, p-JNK1 and p-p38 indicates that cluster had staining of phosphotyrosine, phospho-JNK and phospho-p38, respectively, above background.

which is consistent with their expression of the TRAIL receptor 2, an apoptosis-inducing protein¹⁰. In another cluster (2F7), the plasma membranes of adjoining cells were in close contact to each other (Fig. 3b). This cluster expressed CD36, a cell-surface protein with functions in cell–cell recognition and adhesion¹¹. Other cell phenotypes that are consistent with the potential functions of the cDNAs expressed in the clusters were also present on the microarray (Table 1). Thus, screens using simple detection methods (phosphospecific antibodies and visual inspection) revealed diverse cellular phenotypes in the transfected cell microarray and identified both known and new roles for proteins.

The subcellular distribution of a protein can provide crucial insights to its cellular functions, and its determination is an important part of characterizing a gene of unknown function. Therefore, we also examined, through visual inspection, the subcellular distribution of the proteins expressed in the cell microarrays made with the cDNA collection. We identified 22 clusters expressing proteins with distinct, non-cytoplasmic subcellular distributions (Table 1). Several proteins that are known transcription factors, including ETR101, MEF2C and SP1-like transcription factor, were mainly located in the cell nucleus. This was also true for other proteins with potential nuclear roles, such as deoxycytidine kinase, 3pK (chromosome 3p kinase) and TIP60, as well as others, like phosphatase 1- β (PP1 β), whose subcellular distribution has not been ascertained previously. The location of PP1 β in the nucleus was unexpected and differs from that of its close homologue, PP1 γ , which is also expressed on the array but was found exclusively in the cytoplasm (data not shown).

Transfected cell microarrays have distinct advantages over conventional expression cloning strategies using fluorescence-activated cell sorting or sib selection¹². First, cDNAs do not need to be isolated from the cells showing the phenotype of interest. This significantly accelerates the pace of expression cloning and allows for screens using a variety of detection methods (including those incompatible with cell viability) such as autoradiography and *in situ* hybridization. Our experiments took days to perform instead of the weeks or months necessary with other expression cloning strategies. Second, as the signal is concentrated in a well defined small area in a

transfected cell microarray, cDNAs can be identified whose expression produces only small increases in signal over background. This capacity is well demonstrated by our detection of the receptor for FK506 in host cells with high levels of endogenous FK506-binding proteins. The relatively small increase in binding of the drug over background caused by the overexpression of FKBP12 would have been difficult to detect in the early rounds of sib selection. Third, transfected cell microarrays can be used to screen live cells, allowing the detection of transient phenotypes, such as a change in the intracellular calcium concentration or the movement of a GFP-tagged protein. Fourth, being compact and easy to handle, transfected cell microarrays have economies of scale. Because the microarrays are printed with the same robotic arrayers as traditional DNA arrays, it is feasible to achieve densities of up to 6,000–10,000 cell clusters per standard slide. At these densities, the entire set of human genes could be expressed on a small number of slides. The main difficulty in creating such pan-genomic arrays is the work required in making the cDNA expression constructs to print on the slides. This disadvantage is shared by other microarray-based technologies that use defined sets of cDNAs. Despite the effort required, projects to make collections of full-length cDNAs for all human genes are underway^{13–16} (see also www.hip.harvard.edu/research.html) and will be valuable tools for making the arrays we describe.

Transfected cell microarrays also have certain advantages over protein microarrays, a recently introduced technology in which pure proteins are immobilized on a surface^{17–19}. Although of potentially broad use, highly representational protein microarrays are difficult to make because large numbers of individually purified proteins are needed. Furthermore, it is unclear for how long the proteins on the arrays will be stable once the array is printed. Transfected cell microarrays can be substituted for protein microarrays for some applications, such as in the identification of small molecule targets. Our approach has the advantage that microarrays of DNA, which are stable for months, can be converted when needed into arrays of mammalian cell clusters expressing the encoded gene products.

We have described arrays in which the transfected plasmids direct gene overexpression. However, as antisense technology improves or other methods emerge for decreasing gene function in mammalian cells, it may be possible to also use transfected cell microarrays to screen for phenotypes caused by loss of gene function. The key to spatially restricting transfection without wells is the immobilization of the plasmid DNA in a gel from which it is only accessible to nearby cells. This concept may be generally applicable to the high-throughput and well-less screening in mammalian cells of other molecules, such as chemical compounds, proteins or oligonucleotides. □

Methods

Plasmids

EGFP (enhanced GFP) was expressed from the retroviral vector pBABEpuro²⁰, HA–GST and Myc-tagged FKBP12 from the cauliflower mosaic virus (CMV)-driven prk5 vector²¹, and the human dopamine D1 and serotonin 1A receptors from the LTR (long terminal repeat) driven, bicistronic pMSCV–IRES–EGFP vector. The Genestorm cDNAs are cloned into the CMV-driven pcDNA3.1/GS vector (Invitrogen). These clones were picked at random by Invitrogen from their collection of cDNAs and the clone names are provided in the Supplementary Information. DNA was purified with the Plasmid Maxi or QIAprep 96 Turbo Miniprep kits (Qiagen), and always had an A260/A280 greater than 1.7.

Microarray printing

A robotic arrayer (PixSys 5500; Cartesian Technologies) equipped with stealth pins (ArrayIt SMP4; Telechem) was used to print a plasmid DNA/gelatin solution contained in a 384-well plate onto CMT GAPS glass slides (2549, Corning). The pins deposited about 1-nl volumes 400 μ m apart using a 25–50-ms pin-down-slide time in a 55% relative humidity environment. Printed slides can be stored at 4 °C or at room temperature in a vacuum desiccator. Storage for long periods of time (>3 months) has not appreciably affected performance. Preparation of aqueous gelatin solution is important and is as follows. 0.2% gelatin (w/v) (G-9391; Sigma) was dissolved in MilliQ water by heating and gentle swirling in a 60 °C water bath for 15 min. The solution was cooled slowly to room

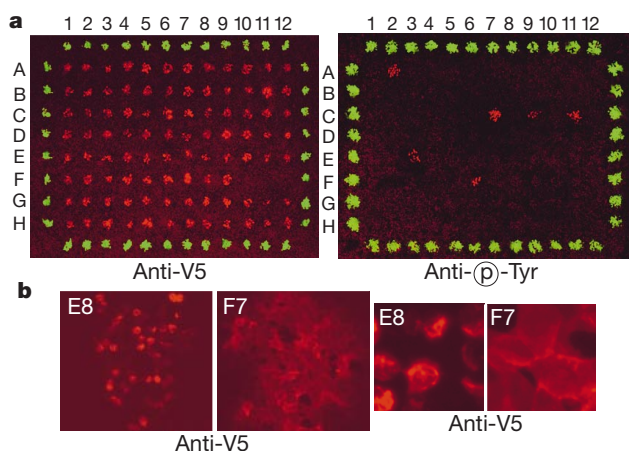


Figure 3 Identification of gene products inducing cellular phenotypes of interest. **a**, Detecting components of tyrosine kinase signalling cascades. V5-epitope-tagged cDNAs (192) in expression vectors were printed in two 8 × 12 subgrids named grid 1 and 2. For ease of determining the coordinates of cell clusters within the arrays, a border around each array was printed with a GFP expression construct. After transfection, separate slides were processed for anti-V5 or anti-phosphotyrosine (anti-p-Tyr) immunofluorescence and Cy3 and GFP fluorescence was detected. Merged images of grid 1 show location of clusters expressing V5-tagged proteins (left panel) and having increased level of phosphotyrosine (right panel). In this particular array, no DNA was printed at coordinates F10–12. **b**, Two examples of the morphological phenotypes detectable in the transfected cell microarrays described in **a**. Clusters shown are E8 and F7 from grid 2.

temperature and filtered through a 0.45- μm cellular acetate membrane and stored at 4 °C. We diluted concentrated solutions of DNA in the gelatin solution to keep the gelatin concentration $>0.17\%$. Unless otherwise specified, final plasmid DNA concentrations were $0.033\ \mu\text{g}\ \mu\text{L}^{-1}$. For the Genestorm clones, the final DNA concentration was $0.040\ \mu\text{g}\ \mu\text{L}^{-1}$.

Reverse transfection of microarrays

We used the Effectene transfection kit (301425; Qiagen) as follows. In a 1.5-ml micro-centrifuge tube, 16 μL enhancer was added to 150 μL EC buffer, mixed, and pre-incubated for 5 min at room temperature. 25 μL Effectene lipid was added, mixed and the entire volume pipetted onto a $40 \times 20\text{-mm}$ cover well (PC200; Grace Bio-Labs). A slide with the printed side down was placed on the cover well such that the solution covered the entire arrayed area while also creating an airtight seal. After a 10–20-min incubation, the cover well was prised off the slide with forceps and the transfection reagent removed carefully by vacuum aspiration. The slide was placed printed side up in a $100 \times 100 \times 10\text{-mm}$ square tissue culture dish and 1×10^7 actively growing HEK293T cells in 25 ml medium (DMEM with 10% IFCS (inactivated fetal calf serum), 50 units mL^{-1} penicillin and $50\ \mu\text{g}\ \text{mL}^{-1}$ streptomycin) were poured into the dish. Three slides can be transfected side by side in this fashion. The cells grew on the slide for 40 h before fixing for 20 min at room temperature in 3.7% paraformaldehyde/4.0% sucrose in PBS. We have also tested other commonly used mammalian cells lines, such as HeLa and A549 cells, and obtained similar results but with transfection efficiencies of 30–50% of those obtained with HEK293T cells.

Immunofluorescence

For immunofluorescence staining, the cells were fixed as above, permeabilized in 0.1% Triton X-100 in PBS for 15 min and probed with primary and secondary antibodies as described²¹. Primary mouse monoclonal antibodies (500 μL per slide) were used at the following concentrations: 1:500 anti-HA ascites fluid (BaBCo), 2 $\mu\text{g}\ \text{mL}^{-1}$ anti-Myc 9E-10 (Calbiochem), 2 $\mu\text{g}\ \text{mL}^{-1}$ anti-V5 (Invitrogen), and 10 $\mu\text{g}\ \text{mL}^{-1}$ 4G10 anti-phosphotyrosine (Upstate Biotechnologies). Primary rabbit polyclonal antibodies (500 μL per slide) were used at the following concentrations: 0.7 $\mu\text{g}\ \text{mL}^{-1}$ anti-p38 pTGpY (Promega), 0.7 $\mu\text{g}\ \text{mL}^{-1}$ anti-JNK pTPpY (Promega), and 0.5 $\mu\text{g}\ \text{mL}^{-1}$ anti-ERK1/2pTEpY (NEB). Secondary antibodies were Cy3-labelled anti-mouse or anti-rabbit antibodies produced in donkeys (Jackson ImmunoResearch) and were used at $3.1\ \mu\text{g}\ \text{mL}^{-1}$.

Laser scanning and fluorescence microscopy

We imaged microarrays at a resolution of 5 μm with a laser fluorescence scanner (ScanArray 5000; GSI Lumonics). GFP and Cy3 emission was measured separately after sequential excitation of the two fluorophores. To obtain images at cellular resolution, cells were photographed with a conventional fluorescence microscope. To measure fluorescence intensity, cell clusters were photographed and the signal intensity quantified with Image Quant (Fuji). All images were pseudocoloured and superimposed using Photoshop 5.5 (Adobe Systems).

FK506 and SCH23390 autoradiography

We added 5 nM dihydro-FK506 [propyl-³H] (NEN) to the cell medium containing the microarrays for 1 h before rinsing once with PBS at room temperature and fixation. When used, rapamycin was added at 1 μM for 30 min before FK506 addition. Before autoradiography, slides were processed for anti-Myc immunofluorescence, scanned at 5- μm resolution and photographed using a fluorescence microscope. We then exposed slides to tritium-sensitive film (Hyperfilm, Amersham) for 4 days. Autoradiographical emulsion was performed as described by the manufacturer (EM-1; Amersham). Image Quant (Fuji) was used to quantify 0.01- mm^2 areas of autoradiograms. For the dopamine receptor experiments, 1 nM SCH23390 [N-methyl-³H] (NEN) was added to the cell medium containing the microarrays for 1 h before rinsing once with PBS at room temperature and fixation. Before processing for autoradiography as above, we photographed GFP fluorescence using a fluorescence microscope.

Received 13 November 2000; accepted 26 March 2001.

- Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
- Lockhard, D. J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**, 1675–1678 (1996).
- Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Kino, T. *et al.* FK-506, a novel immunosuppressant isolated from a Streptomyces. II. Immunosuppressive effect of FK-506 *in vitro*. *J. Antibiot. (Tokyo)* **40**, 1256–1265 (1987).
- Harding, M. W., Galat, A., Uehling, D. E. & Schreiber, S. L. A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. *Nature* **341**, 758–760 (1989).
- Siekierka, J. J., Hung, S. H., Poe, M., Lin, C. S. & Sigal, N. H. A cytosolic binding protein for the immunosuppressant FK506 has peptidyl-prolyl isomerase activity but is distinct from cyclophilin. *Nature* **341**, 755–757 (1989).
- Schreiber, S. L. Chemistry and biology of the immunophilins and their immunosuppressive ligands. *Science* **251**, 283–287 (1991).
- Jacinto, E., Werlen, G. & Karin, M. Cooperation between Syk and Rac1 leads to synergistic JNK activation in T lymphocytes. *Immunity* **8**, 31–41 (1998).
- Gavrieli, Y., Sherman, Y. & Ben-Sasson, S. A. Identification of programmed cell death *in situ* via specific labeling of nuclear DNA fragmentation. *J. Cell Biol.* **119**, 493–501 (1992).
- Walczak, H. *et al.* TRAIL-R2: a novel apoptosis-mediating receptor for TRAIL. *EMBO J.* **16**, 5386–5397 (1997).

- Tandon, N. N., Kralisz, U. & Jamieson, G. A. Identification of glycoprotein IV (CD36) as a primary receptor for platelet-collagen adhesion. *J. Biol. Chem.* **264**, 7576–7583 (1989).
- Simonsen, H. & Lodish, H. F. Cloning by function: expression cloning in mammalian cells. *Trends Pharmacol. Sci.* **15**, 437–441 (1994).
- Strausberg, R. L., Feingold, E. A., Klausner, R. D. & Collins, F. S. The mammalian gene collection. *Science* **286**, 455–457 (1999).
- Kawai, J. *et al.* Functional annotation of a full-length mouse cDNA collection. *Nature* **409**, 685–690 (2001).
- Yudate, H. T. *et al.* HUNT: launch of a full-length cDNA database from the Helix Research Institute. *Nucleic Acids Res.* **29**, 185–188 (2001).
- Wiemann, S. *et al.* Toward a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs. *Genome Res.* **11**, 422–435 (2001).
- MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760–1763 (2000).
- Lueking, A. *et al.* Protein microarrays for gene expression and antibody screening. *Anal. Biochem.* **270**, 103–111 (1999).
- de Wildt, R. M., Mundy, C. R., Gorick, B. D. & Tomlinson, I. M. Antibody arrays for high-throughput screening of antibody–antigen interactions. *Nature Biotechnol.* **9**, 989–994 (2000).
- Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* **18**, 3587–3596 (1990).
- Sabatini, D. M. *et al.* Interaction of RAFT1 with gephyrin required for rapamycin-sensitive signaling. *Science* **284**, 1161–1164 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank C. Wilson and T. Volkert for advice on use of the microarrayer, and K. McKernan and P. McEwan for the neurotransmitter receptor constructs. We also thank R. Young, N. Hacohen and members of the Sabatini laboratory for their support and suggestions; P. Kim and M. Gerlach for help with the manuscript; and Invitrogen for the full-length clones from the Genestorm collection. This research was supported by the G. Harold and Leila Y. Mathers Charitable Foundation, the Whitehead Institute Fellows Program and Corning. Microscopy work was conducted at the W. M. Keck Foundation Biological Imaging facility at the Whitehead Institute.

Correspondence and requests for materials should be addressed to D.M.S. (e-mail: sabatini@wi.mit.edu).

An aminoacyl tRNA synthetase whose sequence fits into neither of the two known classes

Carme Fàbrega*, Mark A. Farrow*, Biswarup Mukhopadhyay†, Valérie de Crécy-Lagard*, Angel R. Ortiz‡ & Paul Schimmel*

* The Skaggs Institute for Chemical Biology, The Scripps Research Institute, Beckman Center, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

† Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

‡ Department of Physiology & Biophysics, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1218, New York, New York 10029, USA

Aminoacyl transfer RNA synthetases catalyse the first step of protein synthesis and establish the rules of the genetic code through the aminoacylation of tRNAs. There is a distinct synthetase for each of the 20 amino acids and throughout evolution these enzymes have been divided into two classes of ten enzymes each^{1,2}. These classes are defined by the distinct architectures of their active sites, which are associated with specific and universal sequence motifs^{1–5}. Because the synthesis of aminoacyl-tRNAs containing each of the twenty amino acids is a universally conserved, essential reaction, the absence of a recognizable gene for cysteinyl tRNA synthetase in the genomes of Archae such as *Methanococcus jannaschii* and *Methanobacterium thermoautotrophicum*^{6–8} has been difficult to interpret. Here we describe a different cysteinyl-tRNA synthetase from *M. jannaschii* and

Deinococcus radiodurans and its characterization *in vitro* and *in vivo*. This protein lacks the characteristic sequence motifs seen in the more than 700 known members of the two canonical classes of tRNA synthetase and may be of ancient origin. The existence of this protein contrasts with proposals that aminoacylation with cysteine in *M. jannaschii* is an auxiliary function of a canonical prolyl-tRNA synthetase^{9,10}.

The ten class I enzymes have a Rossmann nucleotide-binding fold with highly conserved motifs such as the 11-amino-acid signature sequence that ends in HIGH and, separately, the KMSKS pentapeptide^{3–5}. Class II enzymes have three distinct motifs that are part of a seven-stranded β -structure with three α -helices^{1,2}. All known synthetases fit strictly into one of these two classes, so that they are easily recognizable by sequence homology searches with programs such as BLAST or PSI-BLAST¹¹. For example, all cysteinyl tRNA synthetases (CysRSs) can be identified as close homologues that fall into class I. However, the genomes of some Archaea lack recognizable genes for cysteinyl-tRNA synthetase, even though CysRS activity is present in crude extracts¹².

It has been proposed⁹ that another tRNA synthetase replaces cysteinyl tRNA synthetase in *M. jannaschii*. In particular, it was reported^{9,10} that a canonical class II prolyl-tRNA synthetase (termed a ProCysRS) can catalyse aminoacylation with both cysteine and proline. (Interestingly, *Giardia lamblia* encodes a ProCysRS but nonetheless retains a canonical CysRS¹³). Because the presence of ProCysRS raises questions about how one enzyme can accurately maintain aminoacylation with two stereochemically distinct amino acids and two different tRNAs, we searched for an unconventional gene that might encode cysteinyl tRNA synthetase in *M. jannaschii*. Considering that previous methods had failed, we developed a new computational technique for orthologue annotation.

First, we analysed the genome of *M. jannaschii* by comparison with the well-characterized genomes of *Escherichia coli*, *Bacillus subtilis* and *Haemophilus influenza* (see Methods). The idea was to start with well-annotated genes in *M. jannaschii* (about 350 or 20% of the complete genome) and find their most likely orthologues in the other three organisms. The degree of functional matching was evaluated with the ' μ -score', a measure of sequence similarity between two sequences in two different genomes designed so that the higher the μ -score, the higher the likelihood of orthology (see Methods). Conversely, low μ -scores are mainly associated with protein pairs that have different functions. A cutoff (μ_c) of

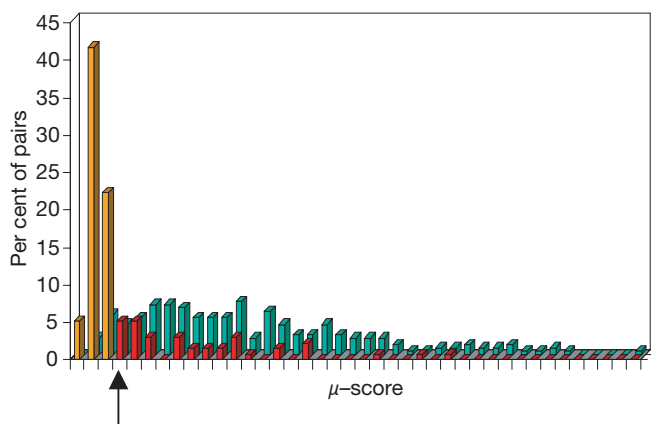


Figure 1 Calibration of the orthologue detection procedure for the *M. jannaschii* genome (see Methods). Pairs of proteins with the same function are represented with green histograms; protein pairs with different functions are in red and orange histograms. The cutoff value of $\mu_c = 2.5$ (arrow) identifies the threshold in the μ -score that can eliminate the bulk of the true negatives (orange bars) while keeping most of the true positives (green bars) in the functional annotation. Complete genome sequences and their functional annotations were downloaded from the European Bioinformatics Institute web site.

about 2.5 was found reliably to separate pairs that had the same function from those that had different functions (Fig. 1). A μ -score of about 2.5 would therefore correspond to a protein pair having on average the lowest possible sequence similarity compatible with keeping the same function in two given genomes. For pairs of distant organisms, this threshold value can potentially relate sequence pairs falling well below the upper limit of the 'twilight zone', the point where sequence similarity is so weak that the trace back to a common ancestor becomes unreliable^{14,15}.

Next, we compared the remaining (about 1,300) genes of *M. jannaschii* with each gene in the whole genomes of the three aforementioned organisms. The idea was to find one or more sequences in *M. jannaschii* that had a μ -score of about $\mu_c = 2.5$ when compared against sequences in the other three organisms. A match of this sort would be a candidate for the missing gene for cysteinyl tRNA synthetase. We found a single match of $\mu = 2.4$ between open reading frame MJ1477 in *M. jannaschii* and the cysteinyl tRNA synthetase in the genome of *B. subtilis* (SYC_BACSU). The MJ1477 coding sequence has 346 amino acids, and the open reading frame did not have significant matches with any other genes in the three genomes.

The gene for MJ1477 (*mj1477*) was cloned and its encoded protein was expressed and purified from *E. coli* (see Methods). We then assayed the CysRS activity of MJ1477 in two ways. Typically, aminoacylation occurs in two steps. In the first step, the amino acid (AA) is condensed with ATP to form the enzyme-bound aminoacyl adenylate (E-AA~AMP). Next, the adenylate reacts with tRNA to form aminoacyl-tRNA (AA-tRNA). In one assay, we monitored the cysteine-dependent exchange of ³²PP_i into ATP and in the second we investigated overall aminoacylation (formation of Cys-tRNA).

The *in vitro* ATP-PP_i exchange assay (Fig. 2a) showed that MJ1477 can activate cysteine in the presence of ATP to yield the corresponding Cys-adenylate (Cys-AMP). The activity is specific for cysteine, because activation does not occur when the isosteric amino acid serine is used instead (Fig. 2a). Activation occurs in the presence as well as in the absence of tRNA. (In contrast, the ProCysRS characterized in ref. 9 activates cysteine only in the presence of tRNA.) We then investigated the specificity of the reaction with respect to the tRNA. *Escherichia coli* tRNA^{Cys} purified by high-performance liquid chromatography (HPLC) was efficiently aminoacylated (Fig. 2b). This aminoacylation was specific for tRNA^{Cys}, because HPLC fractions that were not charged by *E. coli* CysRS were similarly not aminoacylated by MJ1477. Next, and to rule out the possibility of contamination from *E. coli*, we looked at similarities and differences between the *E. coli* and *M. jannaschii* proteins (Fig. 2c). *Escherichia coli* CysRS has some activity with *M. jannaschii* tRNA at 37 °C (considerably less than with its natural substrate), but cannot aminoacylate *M. jannaschii* tRNA at 70 °C. In contrast, purified MJ1477 can aminoacylate *M. jannaschii* tRNA at both 37 and 70 °C (as might be expected for an enzyme from a thermophile). Together, these results strongly support the proposal that MJ1477 is the enzyme catalysing the synthesis of Cys-tRNA^{Cys}.

Attempts to demonstrate the activity of MJ1477 in *E. coli* by complementation of UQ818 cells (ref. 16) containing a temperature-sensitive allele of CysS were unsuccessful because of the toxicity of the expressed protein in the test strain (data not shown). To investigate the activity of MJ1477 further, we determined the kinetic parameters for the activation of cysteine and for the aminoacylation of *M. jannaschii* tRNA. We assayed the rate of synthesis of Cys-AMP by ATP-PP_i exchange (at 70 °C) to give a catalytic constant (k_{cat}) of 0.96 s⁻¹ and a Michaelis constant (K_m) (for cysteine) of 125 μ M. (The calculated k_{cat} for aminoacylation at 70 °C under the conditions used was greater than 0.09 s⁻¹.) These values compare with reported initial rates of 0.132 s⁻¹ for Cys activation and 0.0055 s⁻¹ for aminoacylation (at 70 °C) with the ProCysRS from the same organism¹⁰.

We used the protein-coding sequence of MJ1477 to search for

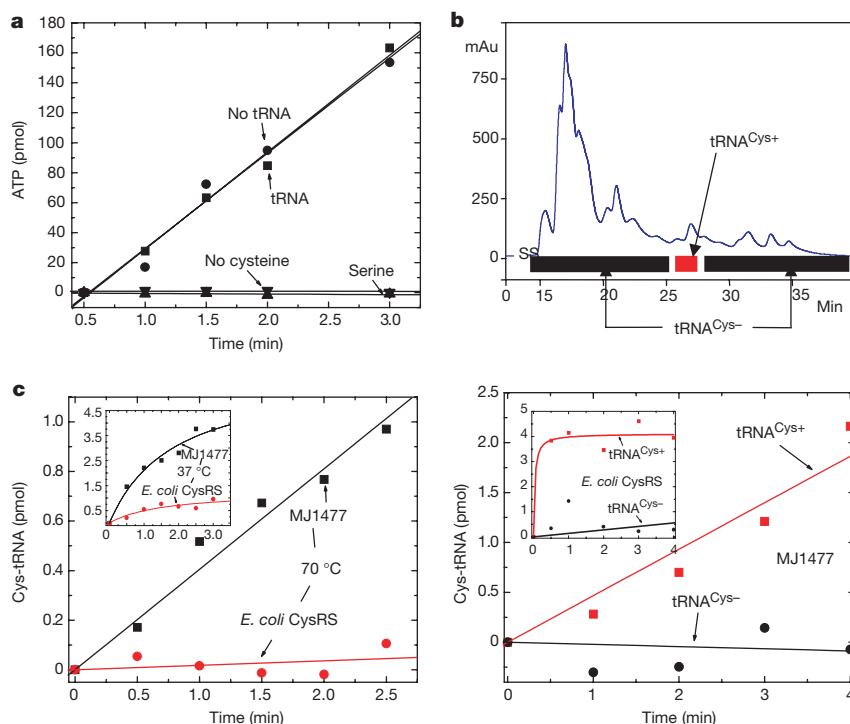


Figure 2 Experimental characterization of the enzymatic activity of MJ1477. **a**, Specific activation of Cys by *M. jannaschii* CysRS at pH 7.0, 70 °C in the presence of 10 mM cysteine and 56 μ M *M. jannaschii* tRNA (squares); in the absence of *M. jannaschii* tRNA (circles); in the presence of 10 mM serine (up triangles; no enzymatic activity detected); and in presence of the enzyme alone (down triangles; no enzymatic activity detected). **b**, Top, HPLC purification of *E. coli* tRNA^{Cys}, with identification of fractions charged with cysteine by *E. coli* CysRS. Bottom, aminoacylation by MJ1477 or *E. coli* CysRS (insert) of

E. coli tRNA^{Cys} (tRNA^{Cys+}) (squares) and *E. coli* tRNA which had been depleted of tRNA^{Cys} (tRNA^{Cys-}) (circles). **c**, Comparison of aminoacylation activity of *E. coli* and *M. jannaschii* CysRS with natural, modified *M. jannaschii* tRNA (56 μ M) at 70 °C. Aminoacylation of *M. jannaschii* tRNA^{Cys} was performed in the presence of MJ1477 (squares) or of *E. coli* CysRS (circles). Insert, same experiment at 37 °C, in the presence of MJ1477 (squares) or *E. coli* CysRS (circles).

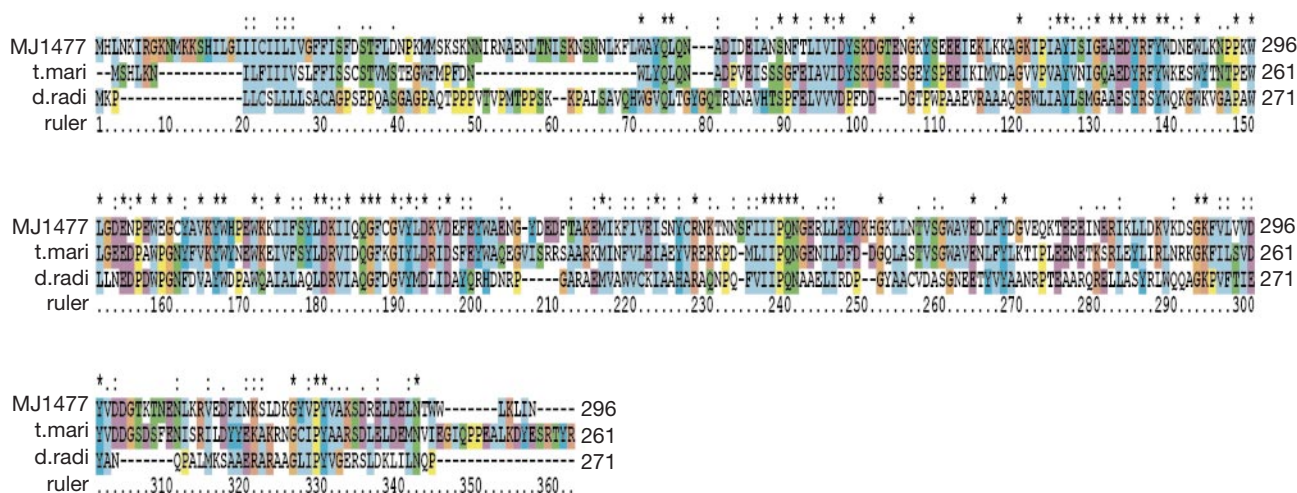


Figure 3 Multiple sequence alignment of MJ1477 with the close homologues detected by PSI-BLAST¹¹. The alignment was obtained with CLUSTALW¹⁷ using default parameters.

homologous sequences in a non-redundant sequence database using PSI-BLAST¹¹. We found two statistically significant hits in the genomes of the eubacteria *Thermotoga maritima* (a thermophile) and *D. radiodurans* (a mesophile). We carried out a multiple sequence alignment (MSA) of the three sequences using the program CLUSTALW¹⁷ with default parameters (Fig. 3). (From this alignment, MJ1477 and the *T. maritima* enzyme appear to be the most similar in sequence of the three apparent homologues.) Using

No class I or class II tRNA synthetases were detected during the database search (see text). Colours follow CLUSTALW convention¹⁷. t.mari, *T. maritima*; d.radi, *D. radiodurans*.

polymerase chain reaction (PCR) and a sample of genomic DNA from *D. radiodurans*, we cloned the MJ1477 homologue from *D. radiodurans* (designated DR0705) and tested it for complementation of *E. coli* strain UQ818. Complementation of the temperature-sensitive lethal *CysS* mutation was observed (Fig. 4). Thus, DR0705 is active *in vivo* in protein synthesis as a cysteinyl tRNA synthetase.

The orthologue detection approach using known sequences of

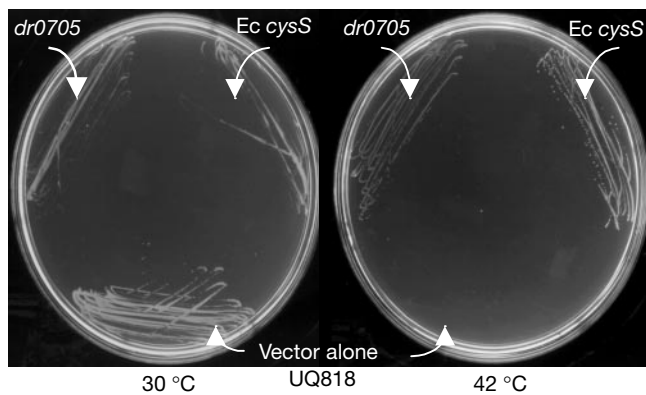


Figure 4 Complementation of the growth defect at 42 °C of *E. coli* strain UQ818. Cells were transformed with vector alone or with plasmids bearing *E. coli cysS* or *D. radiodurans dr0705*, and streaked onto LB plates containing IPTG (1 mM) and ampicillin (100 µg ml⁻¹). Plates were incubated for 24 h at 30 or 42 °C.

class I enzymes has led to the identification of a tRNA synthetase whose sequence bears little or no resemblance to those of known class I enzymes. (It is possible that a three-dimensional structure of MJ1477 or DR0705 would reveal an overall fold similar to that of class I enzymes, but with substantial changes in the active site.) The average per cent identity of MJ1477, for example, with the canonical CysRSs is 17% (not shown), falling well below the twilight zone, conventionally situated in the region of 25–30%¹⁴. This percentage of identity (17%) is significantly lower than the 38% pairwise identity between sequences of canonical CysRSs (not shown). Thus, the alignments provide no support for MJ1477 or its homologues being class I enzymes. More importantly, the hallmarks of class I tRNA synthetases^{3–5}, including the 11-amino-acid signature sequence ending in HIGH and the KMSKS pentapeptide, are missing from the coding sequences of the new CysRSs. These enzymes also lack signatures of the known class II enzymes. Therefore, by the conventional criteria of sequence, MJ1477 and its homologues are the first synthetases to fit into neither of the two known classes of these enzymes.

Thermotoga maritima and *D. radiodurans* present both types of cysteinyl-tRNA synthetase. Small-subunit ribosomal RNA phylogeny has placed *Thermotoga*, together with *Aquifex*, as the deepest and most slowly evolving branch in the tree of life¹⁸. It is possible that the canonical CysRS was acquired later in evolution, and the slow evolutionary pace of *Thermotoga* is the reason why the organism still has both copies. It is also noteworthy that archae such as *M. thermoautotrophicum* lack the novel CysRS. Perhaps the newly identified CysRS is an ancient relic that was gradually eliminated from most organisms. Regardless of the detailed explanation that will eventually emerge, MJ1477 and its homologues are, to our knowledge, the first tRNA synthetases with sequences distinct from those of every other synthetase described. □

Methods

Computational method for orthologue annotation

We used the following protocol to develop the annotation method. (1) Selection of a training set of well-annotated sequences in *M. jannaschii*, *E. coli*, *B. subtilis* and *H. influenzae*. As described in the text, we used about 350 genes, or about 20% of the genome of *M. jannaschii*. (2) For each gene in *M. jannaschii*, the most likely orthologue in the other three genomes was assigned to the sequence with the highest μ -score (see below). (3) The matches obtained were classified into true positives (same function) and true negatives (different function). (4) The cutoff value in the μ -score (μ_c) that optimizes discrimination of true positives versus true negatives for this training set was determined (Fig. 1) using discriminant analysis¹⁹.

The μ -score is obtained as follows: during each of the comparisons described above, sequences were aligned using the Needleman and Wunsch algorithm²⁰ with zero end gaps, using a normalized Gonnet matrix and a threshold ensuring structural similarity²¹. From the sequence alignment score, we developed an orthology likelihood score (μ -score) on the

basis of two premises: first, for two proteins (in two different organisms) to have an orthological relationship they have to be at least structurally related (in addition to having amino acids dictating the same fold, they should have an additional set of similar residues that allows them to carry out a similar function); and second, we assumed the orthological relationship to be biunivocal (only one pair of orthologues exists between two different organisms). With these premises, we defined the μ -score simply as the maximum value of the σ_A score that a protein in one organism can have when compared with all proteins in another organism. In turn, the σ_A score is defined as the number of times the sequence similarity between the two sequences exceeds the minimum value that ensures, on average, a statistically significant structural similarity between both sequences, as derived from training sets of sequence-structure matches²¹. The 'excess' of sequence similarity (over the minimum that provides structural similarity²¹) needed for the pair to have the same function can then be obtained by discriminant analysis (as described above) to yield μ_c .

Cloning and expression of MJ1477 and DR0705

We obtained a clone of the putative CysRS gene from *M. jannaschii* (AMJEC64) (encoding MJ1477) from the American Type Culture Collection (ATCC). *mj1477* was cloned by PCR; for expression and purification of MJ1477, it was subcloned into the maltose-binding protein (MBP) modified expression vector pGEX (Pharmacia) to generate plasmid pMMJ1477. *Escherichia coli* TG1²² was transformed with pMMJ1477, and grown to mid-logarithmic phase at 37 °C in 1 l of Lennox broth containing 100 µg ml⁻¹ ampicillin. The cells were then induced by the addition of 0.5 mM isopropyl 1-thio- β -D-galactopyranoside and incubated for an additional 90 min. Following cell lysis the fusion protein was cleaved using factor Xa and the enzyme purified by a modification of established procedures²³. We determined enzyme concentrations by active site titration²⁴.

We used similar procedures for cloning by PCR of the putative CysRS (Fig. 3) from *D. radiodurans*, using a sample of chromosomal DNA (ATCC13939) from ATCC. The PCR product was cloned into plasmid pTRC99A²⁵ to give plasmid pVDC564, where the *D. radiodurans* gene is under control of an IPTG-inducible promoter. We confirmed the primary structure of the cloned gene (*dr0705*) by direct sequencing methods.

Complementation analysis

Complementation at 42 °C of the growth defect of the temperature-sensitive CysS allele of *E. coli* strain UQ818 (ref. 16) (*lacZ4 gyrA222(nal^R) aroE24 metB rpoB(rif^R) cysS818^{ts}*) was done by standard procedures on LB/ampicillin plates with 1 mM IPTG, using UQ818 cells transformed with the vector (pTRC99A) alone, the vector bearing *dr0705* (pVDC564) and a positive control with the gene for *E. coli* CysRS cloned into a pUC18 vector (gift from G. Eriani).

Enzyme assays and materials

Methanococcus jannaschii tRNA was obtained by phenol extraction from *M. jannaschii* cells²⁶, followed by ion-exchange chromatography (Nucleobond) and desalted with a Centricon 10 concentrator. Total *E. coli* tRNA (Roche Molecular Biochemicals) was purified by HPLC²⁷. Fractions containing tRNA that could be aminoacylated with cysteine (by *E. coli* CysRS) were pooled and desalted using a Centricon-10 concentrator. The remaining tRNA was pooled and desalted. The RNA concentration was determined by absorbance at 260 nm. Cysteinylation synthesis was assessed with the cysteine-dependent ATP-PP_i exchange assay²⁸ at 70 °C in 25 mM HEPES (pH 7.0), 10 mM KF, 25 mM KCl, 3.75 mM MgCl₂, 2 mM dithiothreitol (DTT), 1 mM ATP, 1 mM [³²P]NaPP_i. The kinetic parameters for amino-acid activation were determined using Cys concentrations of 50–2,000 µM. Aminoacylation assays²⁹ were performed in 25 mM Hepes (pH 7.0), 25 mM KCl, 7.5 mM MgCl₂, 2.5 mM DTT, 1 mM ATP, 50 µM [³⁵S]-cysteine, 250 nM MJ1477 or 250 nM *E. coli* CysRS.

Received 31 January; accepted 2 March 2001.

- Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**, 203–206 (1990).
- Cusack, S., Berthet-Colominas, C., Hartlein, M., Nasar, N. & Leberman, R. A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature* **347**, 249–255 (1990).
- Webster, T., Tsai, H., Kula, M., Mackie, G. A. & Schimmel, P. Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase. *Science* **226**, 1315–1317 (1984).
- Hountondji, C., Dessen, P. & Blanquet, S. Sequence similarities among the family of aminoacyl-tRNA synthetases. *Biochimie* **68**, 1071–1078 (1986).
- Ludmerer, S. W. & Schimmel, P. Gene for yeast glutamine tRNA synthetase encodes a large amino-terminal extension and provides a strong confirmation of the signature sequence for a group of the aminoacyl-tRNA synthetases. *J. Biol. Chem.* **262**, 10801–10806 (1987).
- Bult, C. J. et al. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073 (1996).
- Smith, D. R. et al. Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J. Bacteriol.* **179**, 7135–7155 (1997).
- Reeve, J. N. Archaeobacteria then ... Archaea now (are there really no archaeal pathogens?). *J. Bacteriol.* **181**, 3613–3617 (1999).
- Stathopoulos, C. et al. One polypeptide with two aminoacyl-tRNA synthetase activities. *Science* **287**, 479–482 (2000).
- Lipman, R. S., Sowers, K. R. & Hou, Y. M. Synthesis of cysteinyl-tRNA^{Cys} by a genome that lacks the normal cysteine-tRNA synthetase. *Biochemistry* **39**, 7792–7798 (2000).
- Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Hamann, C. S., Sowers, K. R., Lipman, R. S. & Hou, Y. M. An archaeal aminoacyl-tRNA synthetase missing from genomic analysis. *J. Bacteriol.* **181**, 5880–5884 (1999).

13. Bunjun, S. *et al.* A dual-specificity aminoacyl-tRNA synthetase in the deep-rooted eukaryote *Giardia lamblia*. *Proc. Natl Acad. Sci. USA* **97**, 12997–13002 (2000).
14. Chung, S. Y. & Subbiah, S. A structural explanation for the twilight zone of protein sequence homology. *Structure* **4**, 1123–1127 (1996).
15. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.* **12**, 85–94 (1999).
16. Bohman, K. & Isaksson, L. A. Temperature-sensitive mutants in cysteinyl-tRNA ligase of *E. coli* K-12. *Mol. Gen. Genet.* **176**, 53–55 (1979).
17. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
18. Woese, C. R. Interpreting the universal phylogenetic tree. *Proc. Natl Acad. Sci. USA* **97**, 8392–8396 (2000).
19. Johnson, R. A. & Wichern, D. W. *Applied Multivariate Statistical Analysis* (Prentice Hall, Upper Saddle River, New Jersey, 1992).
20. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453 (1970).
21. Abagyan, R. A. & Batalov, S. Do aligned sequences share the same fold? *J. Mol. Biol.* **273**, 355–368 (1997).
22. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
23. Glasfeld, E., Landro, J. A. & Schimmel, P. C-terminal zinc-containing peptide required for RNA recognition by a class I tRNA synthetase. *Biochemistry* **35**, 4139–4145 (1996).
24. Fersht, A. R. *et al.* Active site titration and aminocyl adenylate binding stoichiometry of aminoacyl-tRNA synthetases. *Biochemistry* **14**, 1–4 (1975).
25. Amann, E., Ochs, B. & Abel, K. J. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* **69**, 301–315 (1988).
26. Mukhopadhyay, B., Johnson, E. F. & Wolfe, R. S. Reactor-scale cultivation of the hyperthermophilic methanarchaeon *Methanococcus jannaschii* to high cell densities. *Appl. Environ. Microbiol.* **65**, 5059–5065 (1999).
27. Hale, S. P., Auld, D. S., Schmidt, E. & Schimmel, P. Discrete determinants in transfer RNA for editing and aminoacylation. *Science* **276**, 1250–1252 (1997).
28. Calender, R. & Berg, P. Purification and physical characterization of tyrosyl ribonucleic acid synthetases from *Escherichia coli* and *Bacillus subtilis*. *Biochemistry* **5**, 1681–1690 (1966).
29. Shepard, A., Shiba, K. & Schimmel, P. RNA binding determinant in some class I tRNA synthetases identified by alignment-guided mutagenesis. *Proc. Natl Acad. Sci. USA* **89**, 9964–9968 (1992).

Acknowledgements

This work was supported by grants from the National Institutes of Health (P.S. and to R. S. Wolfe) and the Department of Energy (to R. S. Wolfe) and by a fellowship from the National Foundation for Cancer Research (P.S.). Mount Sinai School of Medicine start-up funds are acknowledged (A.R.O.). We thank Y.-M. Hou for *E. coli* CysRS protein and R. Turner for advice. We thank S. Cusack for reading the manuscript before submission.

Correspondence and requests for materials should be addressed to P.S. (e-mail: schimmel@scripps.edu) or A.R.O. (e-mail: ortiz@inka.mssm.edu).

Developments in analysis

New instrumentation ranges from rugged to portable

TSQ Quantum

ThermoFinnigan www.thermoFinnigan.com

Compact triple-quadrupole mass spectrometer

Use of a 90° collision cell reduces the size of this triple quadrupole mass spectrometer to 55 × 56 × 79 cm. The cell's square quadrupole technology is claimed to raise quantitative performance by optimizing ion transmission and increasing the signal-to-noise ratio. High ion transmission and mass selectivity make for high resolution, minimizing matrix interferences to resolve ambiguities normally associated with low-resolution analysis. Xcalibur control and analysis software is complemented by a suite of software products for quantitative metabolite identification, drug discovery and proteomic applications.

TopCount NXT

Packard Instrument Company

www.packardinstrument.com

On-line calculations for speedier analysis

The TopCount NXT counter now features a number of special calculations to help accelerate and simplify the analysis of results. Specific calculations include per cent of lysis for ⁵¹Cr-release cytotoxicity assays, per cent inhibition for various types of receptor binding assays, per cent of control for determining relative binding, inhibition or lysis, and Log c.p.m. to ease reading of results in which the c.p.m. extend over many decades. Raw counts



Third program: the compact TSQ Quantum.

may also be reported for use in chromatographic analysis software. For complex calculations, data can be provided in ASCII format for direct importation into Microsoft Excel or other software packages for processing. TopCount NXT software can also be used to initialize Excel macros at the end of assay batches or groups, then return control to the microplate counter for more counting.

YSI 8500

Analytical Technologies

www.analyticaltechnologies.co.uk

In-situ CO₂ measurement

The YSI 8500 monitor makes it possible to measure dissolved carbon dioxide *in situ* in bioprocess development applications. Engineered to fit within a variety of bioreactors, the unit provides real-time data on dissolved carbon dioxide levels. High concentrations have

been shown to inhibit cell growth and product formation in mammalian cells and to alter the glycosylation pattern of recombinant proteins. By maintaining low and constant levels, the production rate of pharmaceuticals, proteins and antibodies can be maximized. Suggested uses include critical fermentation and cell culture processes.

D-luciferin

Pierce

www.piercenet.com

Pure and functional

Primarily used in reporter assays, ATP assays and kit manufacturing, Pierce's D-luciferin is supplied as free acid, sodium salt or potassium salt. Stringent specifications maintain purity and chiral-HPLC testing is carried out to ensure that the luciferin is the active form. The product is available in sizes ranging from 5 mg to custom bulk packaging.

High-pressure gas cell

SensIR Technologies

www.sensit.com

FT-IR on the move

This new gas cell accessory is designed to fit the TravelIR portable Fourier transform infrared (FT-IR) spectrometer, the DuraSamplerIR and DuraScope in-compartment FT-IR analysis systems. It is capable of testing gases up to 1,300 p.s.i. and, although compact in size, creates a 10-cm path length. High pressure facilitates handling of small samples directly from the container, producing high quality spectral data with a minimum number of required scans. The internal cell is built using 316 stainless steel, sapphire windows and Kalrez 'O' ring, making it inert to most gases presented for analysis.

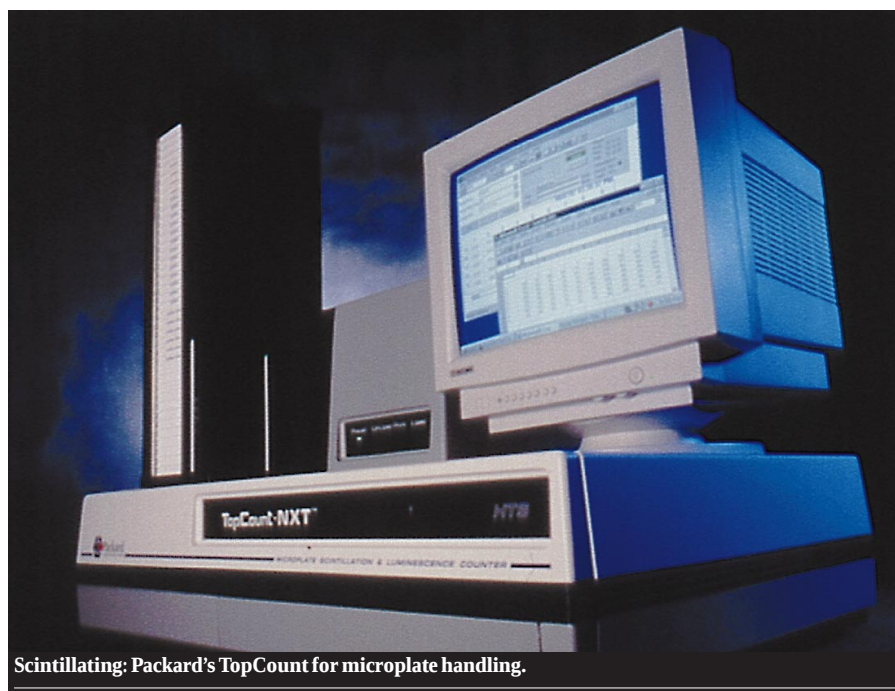
ELAN DRC ICP-MS

PerkinElmer

www.perkinelmer.com

Application note on arsenic detection

This application note describes the use of the ELAN Dynamic Reaction Cell (DRC) inductively coupled plasma mass spectrometry (ICP-MS) system to determine arsenic in chloride matrices. The authors explain that it is difficult to determine low levels of arsenic in chloride matrices with conventional ICP-MS because of the formation of ⁴⁰Ar³⁵Cl⁺ interference in the plasma, and note that recent environmental regulations and the need to detect different organic and inorganic species of arsenic have driven detection limits down. The ELAN DRC ICP-MS system eliminates the ⁴⁰Ar³⁵Cl⁺ background interference using chemical resolution and adjustment of the reaction cell bandpass.



Scintillating: Packard's TopCount for microplate handling.



Concentration: the CentriVap range.

CentriVap

Labconco

www.labconco.com

New brochure features centrifugal concentrators and cold traps

This full-colour brochure contains details of Labconco's redesigned CentriVap centrifugal concentrators, including the CentriVap concentrator and the Ultra-low cold trap. These use a combination of centrifugal force, vacuum and heat to speed evaporation of multiple small samples. A new system incorporates a gel dryer along with the concentrator and cold trap. Evaporation rates and specifications for each product are provided. Accessories, including rotors, a strobe light, vacuum pumps and chemical traps are also described.

AA-6200

Shimadzu

www.shimadzu.com

Automated flame/furnace atomic absorption analyser

The AA-6200 system includes flame and furnace atomic absorption (AA) with true double-beam optics, deuterium background, a PC and the ASC-6100 autosampler, which can be used for both flame and graphite furnace analysis. The AA autosampler can mix any combination of up to five solutions into a homogeneous mixture for furnace analysis. A sample, two matrix modifiers, a diluent and a matrix spike can be combined into a homogeneous mixture. Wizard software containing quality assurance and quality control protocols completes the system.

Agilent 7500c

Agilent

www.agilent.com

Extended ICP-MS analysis capability

The 7500c inductively-coupled plasma mass spectrometer (ICP-MS) is designed for the analysis of samples that cannot be handled using current quadrupole ICP-MS technology. To overcome interference, the system uses an octopole reaction cell. During the analysis, the interfering species are dissociated by a

reaction gas within the cell, eliminating the interference and enabling the analyte to be determined. An octopole reaction system (ORS) employs 'soft' or moderately reactive gases that minimize the potential for complications from side-reactions. The passive ORS requires no optimization of scanning voltages and its modular configuration is easily removed for maintenance. With simultaneous multi-element capability, high sample throughput and part-per-trillion sensitivity, the 7500c is described by the manufacturer as a key analytical tool for laboratories performing trace metals analysis in complex samples.

I-CHEM vials

Nalgene

www.nalgenunc.com

Low-level TOC certified vials

I-Chem 40-ml clear vials are certified for low-level total organic carbon (TOC) analysis, containing less than 50 p.p.b. TOC. They are made out of borosilicate glass with Teflon/silicone septum closures. Each vial is bar-coded with an individual container and lot number that can be read either manually or by laser and that is traceable to the certificate of analysis included with each case of vials. Unattached sample analysis labels are also included. The vials meet all performance-based specifications for volatile organic analysis and are suitable for use in environmental laboratories.

ZQ mass detector

Waters

www.waters.com

Rugged, but sensitive

This single-quadrupole atmospheric pressure ionization (API) mass spectrometry detector requires only 15 inches (39 cm) of linear bench space and is claimed to be both rugged and exceptionally sensitive. Models are available with a mass range of either 2,000 or 4,000 daltons and the unit is optimized for use with Waters' Alliance, Alliance HT and component systems anchored by the Waters 600 solvent delivery module. It will also operate with HPLC systems made by other manufacturers.

Infinityplus

Varian

www.varianinc.com

NMR spectrometer for solid-state applications

Solids nuclear magnetic resonance (NMR) spectrometry has expanded to include the study of biomolecular systems such as receptor proteins, biological membranes and the metallo-organic protein complexes associated with photosynthesis, for example, hemes. The Infinityplus is designed specifically for analysis of these solid-state samples and is also suitable for studies involving more traditional solid-state materials such as polymers, ceramics, and the catalysts used to generate petroleum products.

SilverQuest

Invitrogen

www.invitrogen.com

MS-compatible silver stain

This mass spectrometry-compatible silver-staining kit is designed for maximum sensitivity with minimum protein modification. It can detect as little as 0.3 ng of protein in less than 90 minutes. A non-glutaraldehyde sensitizer and destaining solution are said to provide more efficient trypsin digestion, greater protein coverage and improved mass spectrometry analysis without sacrificing sensitivity of detection.

Genesis AQ

Jones Chromatography

www.jones-chrom.co.uk

4 µm aqueous columns

Genesis AQ columns are suitable for use with 100% aqueous mobile phases without the typical problems of reduced and variable retention times and inferior peak shape. Column equilibration times in water are short and the columns are stable in up to 100% organic mobile phases, making them suitable for a wide range of applications. They are available in a variety of configurations, from microbore to analytical to preparative.

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

Polyclonal Antibodies & Custom Peptides

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

ResGen™

Invitrogen Corporation

U.S. or Canada 800-533-4363

FAX 256-536-9016

Check out our web site at
www.resgen.com

REAGENT KITS

NEW

FLUOROSPARK PTP KIT

Simple, Rapid and Sensitive
Quantifies PTP activity

ESTROGEN RECEPTOR KIT

High throughput screening of endocrine
disrupting chemicals

SLP REAGENT KIT

Highly sensitive peptidoglycan and
(1-3)-β-glucan Assay

Wako

Wako USA

877-714-1921

Wako GmbH

(49)02131 3110

www.wako-chem.co.jp